

Institute for Advanced Simulation (IAS)  
Jülich Supercomputing Centre (JSC)

## **Dichtebestimmung von Menschenmassen unabhängig von der Kameraausrichtung**

*Daniel Salden*





# **Dichtebestimmung von Menschenmassen unabhängig von der Kameraausrichtung**

*Daniel Salden*

Berichte des Forschungszentrums Jülich; 4355  
ISSN 0944-2952  
Institute for Advanced Simulation (IAS)  
Jülich Supercomputing Centre (JSC)  
Jül-4355

Vollständig frei verfügbar im Internet auf dem Jülicher Open Access Server (JUWEL)  
unter <http://www.fz-juelich.de/zb/juwel>

Zu beziehen durch: Forschungszentrum Jülich GmbH · Zentralbibliothek, Verlag  
D-52425 Jülich · Bundesrepublik Deutschland  
☎ 02461 61-5220 · Telefax: 02461 61-6103 · e-mail: [zb-publikation@fz-juelich.de](mailto:zb-publikation@fz-juelich.de)

## Abstract

Analysis and simulation of crowds and people streams are one area of research at the Jülich Supercomputing Centre (JSC) at the Forschungszentrum Jülich. The main focus of the work is on people security. For the safety of people the distribution and the density of the people is most important. This master thesis presents a method to determine the distribution of people independent from the position and viewing of the camera. To estimate the positions of the people, the coordinate systems of the world and the camera have to be related to each other. Therefore in a calibration process the translation and the rotation between these two coordinate systems are calculated with so called point correspondences, which contain 2D image points and their corresponding 3D points inside the world coordinate system. To calculate the positions and thus the density of the people we form a straight line from the point of projection of the camera through the 2D image point in the image plane into the scene in front of the camera. The position of the persons' head is the intersection point between the line and the plane at the average height of the persons. The positions of the people on the ground plane are now the input for building a Voronoi diagram. The density for a person is inversely proportional to its Voronoi cell area. The introduced method is tested at a toy figure model to analyze the quality of the proposed procedure. Also an image of the Loveparade 2010 in Duisburg is inspected to get information about the density of the people for this tragic happening.



## Zusammenfassung

Die Analyse und Simulation von Menschenströmen und -ansammlungen ist ein Forschungsgebiet des Jülich Supercomputing Centre (JSC) im Forschungszentrum Jülich. Der Fokus der Arbeiten liegt auf der Betrachtung von Sicherheitsaspekten. Für die Sicherheit der Personen spielt die Verteilung bzw. die Dichte der Personen eine bedeutende Rolle. Die Masterarbeit liefert ein Verfahren um die Verteilung der Personen auf der Ebene unabhängig von der Ausrichtung der Kamera zu bestimmen. Bevor die Standorte der Personen ermittelt werden können, muss zunächst der Zusammenhang zwischen Weltkoordinaten- und Kamerakoordinatensystem hergestellt werden. Dazu wird in einem Kalibrationsschritt mithilfe von 3D/2D-Punktkorrespondenzen die Rotation und Translation zwischen Welt- und Kamerakoordinatensystem ermittelt. Mithilfe dieser äußeren Orientierung kann nun die Positions- und Dichtebestimmung durchgeführt werden. Als Ausgangspunkt nimmt man die 2D-Bildkoordinaten der Personenköpfe im Bild. Ausgehend vom Projektionsursprung bildet man nun Geraden durch die 2D-Bildpunkte in die 3D-Szene. Die Schnittpunkte dieser Geraden mit der Ebene, welche auf Höhe der Personenköpfe liegt, sind schließlich die 3D-Positionen der betrachteten Köpfe. Die Standortpositionen auf dem Boden werden schließlich als Ausgangspunkte für die Bildung eines Voronoi-Diagramms verwendet. Die einzelnen Voronoi-Zellen stellen dabei den persönlichen Raum einer jeden Person dar. Die Dichte für die einzelne Person ist dabei umgekehrt proportional zur Fläche seiner Voronoi-Zelle. Als Anwendung des Verfahrens werden ein Modell mit Spielzeugfiguren und eine Aufnahme der Loveparade untersucht. Dabei wird insbesondere die Abhängigkeit zwischen der angenommenen Personengröße und der Dichte sowie die Güte des Verfahrens analysiert.





# Inhaltsverzeichnis

## Kurzfassung

1	Einleitung	1
1.1	Motivation . . . . .	1
1.2	Aufgabenstellung . . . . .	2
2	Grundlagen	5
2.1	Koordinatensysteme . . . . .	5
2.1.1	Weltkoordinatensystem . . . . .	5
2.1.2	Kamerakordinatensystem . . . . .	7
2.1.3	Bildkoordinatensystem . . . . .	7
2.2	Koordinatentransformationen . . . . .	8
2.3	Lochkameramodell . . . . .	10
2.4	Projektion . . . . .	11
2.4.1	Mathematischer Hintergrund . . . . .	13
2.5	Kameraparameter . . . . .	18
3	Kamera Kalibrierung	21
3.1	Perspective-n-Point Problem (PnP-Problem) . . . . .	23
3.1.1	POSIT . . . . .	23
4	Implementierung	31
4.1	Verfahrensbeschreibung . . . . .	31
4.2	Kalibrierung . . . . .	33
4.3	Positionsberechnung . . . . .	33
4.4	Dichtebestimmung . . . . .	36
4.5	Dichtevisualisierung . . . . .	36
4.6	Grafische Oberfläche . . . . .	38

---

5	Ergebnisse und Bewertung	41
5.1	Anwendung im Modell . . . . .	41
5.2	Anwendung an realem Szenario . . . . .	45
5.2.1	Fahrrad-Experiment . . . . .	49
5.3	Fehlerbetrachtung . . . . .	52
6	Zusammenfassung	59
6.1	Ausblick . . . . .	60
A	Anhang	63
A.1	Modellaufbau . . . . .	63
A.2	Benutzungsanleitung . . . . .	64
A.3	Ergebnisse verschiedener angenommener Größen . . . . .	67
A.4	OpenCV . . . . .	71
A.5	Verwendete Hilfsmittel . . . . .	71





# 1 Einleitung

## 1.1 Motivation

In den letzten Jahren erleben Großveranstaltungen, wie das Public Viewing zur Fussball-Europameisterschaft oder das Besuchen von großen Konzerten, steigende Beliebtheit. Oft sind mehrere zehntausend Besucher auf engstem Raum versammelt. Damit solch eine Veranstaltung reibungslos und sicher stattfinden kann, müssen im Vorfeld bereits Rettungswege und Notausgänge bereitgestellt werden. Aber auch während der Veranstaltung ist es enorm wichtig, die Menschenmassen im Auge zu haben, um erste Anzeichen für eine Überfüllung rechtzeitig wahrzunehmen und entsprechend gegen wirken zu können. Für die Veranstalter und Sicherheitskräfte einer Großveranstaltung ist es wichtig, neben der absoluten Zahl der Besucher auch die Verteilung der Besucher zu kennen. Denn oft führt eine ungleiche Verteilung zu großen Sicherheitsrisiken für die Besucher.

In der Vergangenheit hat sich gezeigt, dass es aufgrund von überfüllten Veranstaltungsorten oft zu Unfällen auf Massenveranstaltungen kommt. Um solch eine Katastrophe wie z.B. 2010 bei der Loveparade in Duisburg zu verhindern, wird sich intensiv mit der Fragestellung befasst: Wie kann man gefährlich hohe Personendichten bereits im Vorfeld ausfindig machen und dementsprechend entgegen wirken? Auch im Forschungszentrum Jülich arbeitet man an dieser Fragestellung. Das Projekt „BaSiGo - Bausteine für die Sicherheit von Großveranstaltungen“ befasst sich mit dem Ziel, Großveranstaltungen sicherer zu machen [1]. Das Forschungszentrum Jülich beschäftigt sich dabei unter anderem mit der Erfassung von Menschenmengen in realen und künstlichen Umgebungen [2].

Ein Ziel dieser Arbeit ist es, den Ist-Zustand der Personenverteilung bei einer Menschenansammlung zu ermitteln. Dazu soll anhand einer Kamera-

aufnahme, in der die Personen markiert werden, die Dichteverteilung der Personen bestimmt werden. Die Dichte wird dabei durch die Anzahl der Personen pro Quadratmeter beschrieben. Doch um die Dichte berechnen zu können, benötigt man die Standortpositionen der jeweiligen Personen. Die Pixelkoordinaten bzw. Bildinformationen geben für das menschliche Auge einen ersten Eindruck von der Verteilung der Personen, sind jedoch für die genaue Standortbestimmung nicht direkt verwertbar. Die Projektion der 3D-Szene auf die Bildebene gibt dem Betrachter ein perspektivisch transformiertes Bild, in dem weiter entfernte Personen kleiner, und zum Betrachter nähere Personen größer abgebildet werden. Aus den Pixelinformationen alleine kann man so zunächst keine Rückschlüsse auf die 3D-Geometrie ziehen. Genau diese 3D-Geometrie wird jedoch benötigt, um die Positionskoordinaten der Personen zu erhalten.

## 1.2 Aufgabenstellung

Bei der Projektion des Raumes auf die Bildebene wird die Tiefeninformation der einzelnen Bildpunkte verworfen. Somit gibt es anschließend keine Informationen mehr darüber, wie weit das dargestellte Pixel bei der Aufnahme von der Kamera entfernt war. Man hat lediglich die Pixelkoordinaten der Personen bzw. - in der Regel bei dichten Menschenmengen - nur die Koordinaten der Personenköpfe. Man weiß jedoch, dass ein Pixel einen Punkt repräsentiert, der auf der Geraden liegt, die durch den Projektionsursprung und dem Pixel in der Bildebene geht.

Die Aufgabe der Masterarbeit besteht nun daraus, die 3D-Positionen der markierten Bildpunkte im Kamerakoordinatensystem anzugeben. Anhand der ermittelten Positionen soll anschließend ein Wert für die Dichte jeder einzelnen Person bestimmt werden. In einem Kalibrationsschritt soll dazu mithilfe von Punktkorrespondenzen, welche zu 2D-Bildpunkten die zugehörigen 3D-Weltkoordinaten angeben, die äußere Orientierung der Kamera bestimmt werden. Neben den Punktkorrespondenzen werden auch die Werte der Kameramatrix, welche die interne Geometrie der Kamera, wie z.B. die Brennweite und den Kamerahauptpunkt, genauer beschreiben, benötigt. Durch den Kalibrationsschritt erhält man die

Lage und Ausrichtung der Kamera zur 3D-Welt. Mit dieser Information kann man nun Aussagen über den Verlauf einer Geraden im Raum treffen. Zur Bestimmung der genauen Position des Punktes auf der Geraden benötigt man jedoch eine weitere 3D-Angabe zum Bildpunkt. Dazu wird davon ausgegangen, dass sich die betrachteten Personen alle auf einer ebenen Fläche befinden und die Personengrößen nur gering voneinander abweichen. Man definiert eine Ebene parallel zur Standfläche auf Höhe der Durchschnittsgröße der Personen. Die genaue Position auf der Geraden soll nun durch eine Schnittpunktberechnung zwischen der Geraden und der parallel zum Boden verlaufenden Ebene ermittelt werden. Der Schnittpunkt enthält die 3D-Koordinaten für den Kopf bzw. die Markierung der Person. Die Position der Person auf dem Boden erhält man durch die Projektion des berechneten 3D-Punkt orthogonal zum Boden.

Anschließend soll anhand der Positionskoordinaten ein Voronoi-Diagramm erstellt werden. Die Dichte innerhalb einer Zelle wird durch die Größe der Voronoi-Zelle bestimmt. Der Wert der Dichte ist dabei umgekehrt proportional zum Flächeninhalt der Voronoi-Zelle. Das Finden der einzelnen Personen im Bild ist nicht Bestandteil der Arbeit [3, 4]. Für die Arbeit werden daher die 2D-Bildkoordinaten der Personenköpfe vorab interaktiv vom Benutzer eingegeben.

Nach dem in Kapitel 2 die benötigten grundlegenden Begriffe und Verfahren der Bildverarbeitung erläutert werden, befasst sich das 3. Kapitel mit der Kalibration der Kamera durch Kalibrationspunkte. Die Kamerakalibration liefert die äußere und innere Orientierung der Kamera, welche in Kapitel 4 für die softwaretechnische Umsetzung des Verfahrens benötigt wird. Die Ergebnisse der Arbeit werden am Ende der Arbeit an konkreten Anwendungsbeispielen vorgeführt. Dazu wird zum einen ein Modell von Spielzeugfiguren verwendet, bei dem die genauen 3D-Positionen von Kamera und Personen bekannt sind, um etwas über die Güte des Verfahrens aussagen zu können. Ebenso wird ein Bild der Loveparade 2010 betrachtet, bei dem jegliche Informationen zur Kamera- und Personenpositionen unbekannt sind. In einer Fehlerbetrachtung wird anschließend auf die Probleme und Ursachen für Fehler eingegangen, bevor das erarbeitete Verfahren im Schlussteil kurz zusammengefasst und bewertet wird.





## 2 Grundlagen

Im Bereich der Bildverarbeitung gibt es eine Reihe von Modellen und Verfahren, welche im folgenden Kapitel zunächst erläutert werden, bevor sie im weiteren Verlauf der Arbeit zur Positions- und Dichtebestimmung ihre Anwendung finden.

### 2.1 Koordinatensysteme

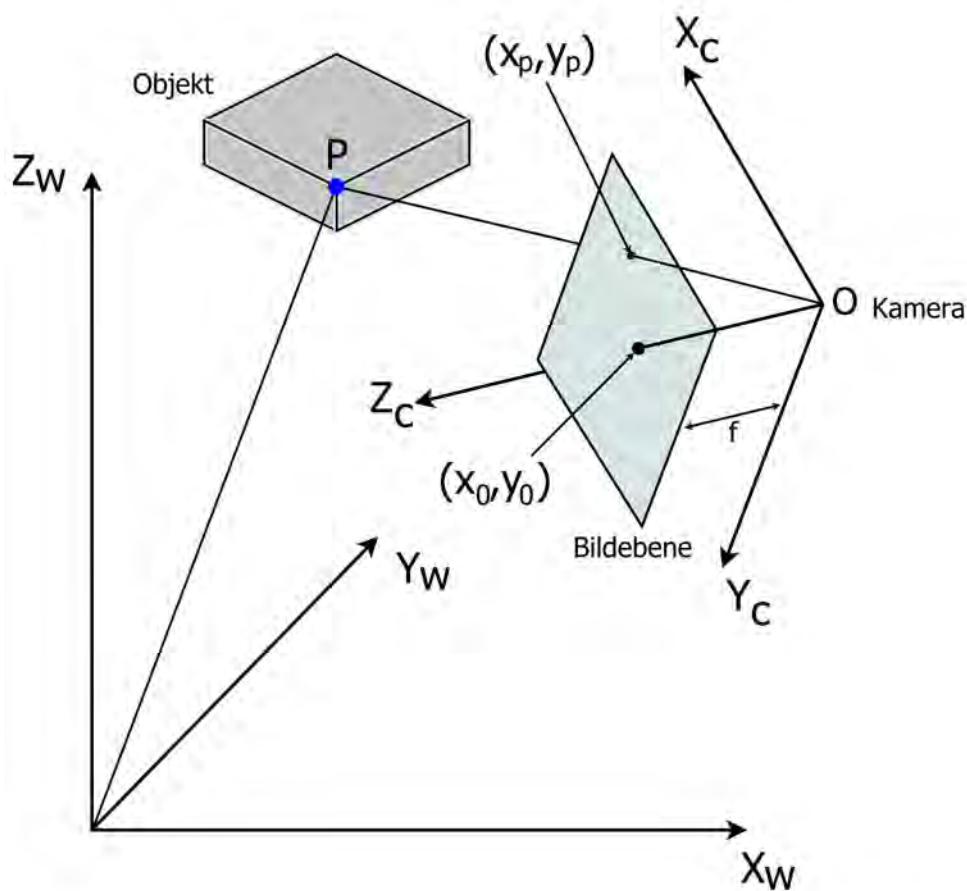
Im Bereich der Bildverarbeitung gibt es verschiedene Koordinatensysteme, welche im 3D-Bereich oder in der 2D-Bildebene definiert sind. Das Weltkoordinatensystem dient dabei als Referenzkoordinatensystem für die jeweiligen relativen Koordinatensysteme. Im Weltkoordinatensystem können alle weiteren relativen Koordinatensysteme in Bezug zueinander gesetzt werden. Dazu müssen die Positionen und Orientierungen der relativen Koordinatensysteme zum Weltkoordinatensystem bekannt sein.

#### 2.1.1 Weltkoordinatensystem

Das Weltkoordinatensystem befindet sich in der realen 3-dimensionalen Szene und dient als Grundlage für alle weiteren relativen Koordinatensysteme. Es wird durch ein kartesisches Koordinatensystem mit orthogonalen Achsen beschrieben. Ein Punkt im Weltkoordinatensystem hat die Koordinaten  $P_w = \begin{pmatrix} x_w \\ y_w \\ z_w \end{pmatrix}$ , wobei das Weltkoordinatensystem üblicherweise als rechtshändiges System definiert ist (vergleiche Abbildung 2.1). Wurde das Koordinatensystem einmal in der 3D-Szene definiert, so wird es in der Regel nicht mehr verschoben. Mithilfe des Weltkoordinatensystems

kann man weitere Objekte in der 3D-Szene platzieren bzw. deren Position beschreiben.

Die Punkte eines Objektes in der 3-dimensionalen Szene werden im weiteren Verlauf in Weltkoordinaten angegeben. Für die spätere Untersuchung des Zusammenhangs zwischen 2D-Bildebene und 3D-Szene spielt die Lage bzw. Position der Kamera eine wichtige Rolle. Die Lage der Kamera wird durch die Angabe der Verschiebung und der Drehung zum Weltkoordinatensystem angegeben.



**Abbildung 2.1:** Zusammenhang zwischen dem Weltkoordinatensystem  $(X_W, Y_W, Z_W)$ , dem Kamerakoordinatensystem  $(X_C, Y_C, Z_C)$  und der Bildebene

### 2.1.2 Kamerakoordinatensystem

Das Kamerakoordinatensystem beschreibt die Sicht und Orientierung der Kamera. Der Ursprung des Kamerakoordinatensystems liegt im Projektionsursprung der Kamera. Ein Punkt im Kamerakoordinatensystem besitzt die Koordinaten  $P_C = \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix}$ . Die Position der Kamera im Weltkoordinatensystem wird im Folgenden mit  $C$  angegeben. Die  $X_C Y_C$ -Ebene des Kamerakoordinatensystems liegt parallel zur Bildebene. Der Vektor  $Z_C$  verläuft entlang der optische Achse der Kamera. Im Vergleich zum Weltkoordinatensystem handelt es sich beim Kamerakoordinatensystem um ein linkshändiges Koordinatensystem. Die Position und Ausrichtung des Kamerakoordinatensystems ist völlig unabhängig vom Weltkoordinatensystem (vergleiche Abbildung 2.1).

### 2.1.3 Bildkoordinatensystem

Hinter dem Bildkoordinatensystem verstecken sich zwei Begriffe, die in diesem Zusammenhang von Bedeutung sind. Zum einen die Bildebene, welche die Ebene bezeichnet, auf der ein 3-dimensionaler Raum durch Projektion abgebildet wird. Desweiteren gibt es noch die Pixelmatrix, welche als Diskretisierung der Bildebene die einzelnen Bildelemente bzw. Pixel enthält.

#### Bildebene

Die Bildebene kann entweder hinter oder vor dem Projektionszentrum  $O$  liegen. Liegt die Ebene hinter dem Projektionszentrum, so werden die Projektionsstrahlen hinter dem Projektionszentrum verlängert und es entsteht ein spiegelverkehrtes Abbild der 3D-Szene. Die Bildebene wird oft auch als Projektionsebene bezeichnet. Auf der Bildebene sind die aufgenommenen Objekte perspektivisch verzerrt sichtbar. Sie liegt parallel zur  $X_C Y_C$ -Ebene des Kamerakoordinatensystems mit dem Abstand  $f$  zum Kamerakoordinatensystemursprung. Der Abstand  $f$  der Bildebene zum Projektionsursprung wird als Brennweite der Kamera bezeichnet und definiert den Sichtbereich der Kamera.

## Pixelmatrix

Die Pixelmatrix beinhaltet entsprechend der Auflösung des Bildes ein zweidimensionales Feld von gleich großen Pixeln. Jedes Pixel ist dabei die Diskretisierung des entsprechenden Bereichs in der Bildebene. Die Indizierung der einzelnen Pixel erfolgt dabei zeilen- und spaltenweise in der oberen linken Ecke beginnend. Ein einzelnes Pixel wird durch den Index der Spalte  $u$  und durch den Index der Zeile  $v$  angegeben  $(u,v)$ .

## 2.2 Koordinatentransformationen

Koordinatentransformationen ermöglichen die Umrechnung von Punkten zwischen zwei Koordinatensystemen. Dafür muss die Lage und Orientierung der zwei Koordinatensysteme zueinander bekannt sein. Die Transformation der Koordinaten erfolgt dabei mithilfe von Transformationsoperationen. Eine Übersicht über die wichtigsten Transformationsoperationen liefert die folgende Auflistung, in der einige Operationen kurz erläutert werden.

**Rotation** Durch die Rotation wird die Drehung eines Objektes realisiert. Jedes Objekt im 3-dimensionalen Raum kann dabei beliebig im Koordinatensystem gedreht werden. Für die spätere Umrechnung vom Weltkoordinatensystem ins Kamerakoordinatensystem wird die Rotation durch eine 3x3-Rotationsmatrix ausgedrückt.

**Translation** Die Translation beschreibt die Verschiebung bzw. die Bewegung eines Objektes entlang eines Vektors  $\vec{t} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}$ . Die Translation spiegelt sich in der Transformationsmatrix in der letzten Spalte wieder.

**Skalierung** Die Skalierung bewirkt eine Veränderung der Achseneinheiten. Mathematisch werden dazu die Koordinaten des zu transformierenden Vektors mit den Skalierungsfaktoren für die  $x$ -,  $y$ - und  $z$ -Achse multipliziert.

**Spiegelung** Die Spiegelung bewirkt, dass eine Achse in die entgegengesetzte Richtung gespiegelt wird. Eine Achse, die vorher in positiver Richtung nach oben zeigte, zeigt anschließend nach unten. Die Spiegelung kann verwendet werden, um aus einem linkshändigen Koordinatensystem ein rechtshändiges Koordinatensystem zu erhalten.

### Homogene Erweiterung

Die Transformationsoperationen können alle entweder als Matrixmultiplikation oder als Vektoraddition ausgedrückt werden. Um alle Operationen jedoch als geschlossene Matrix ausdrücken zu können, behilft man sich anhand von homogenen Koordinaten. Dabei erhalten die Vektoren und Matrizen eine zusätzliche Komponente bzw. Zeile. Hintergrund der zusätzlichen Komponente ist, dass man alle Transformationen (Rotation, Translation, Skalierung, usw.) geschlossen in einer Matrix ausdrücken kann. Würde man auf die homogene Erweiterung verzichten, so müsste man die Translation in einem extra Schritt hinzurechnen:

$$\vec{x}' = R\vec{x} + \vec{t}$$

mit

$$R = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{pmatrix}, \vec{t} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}$$

Entsprechend erweitert man die Transformationsmatrix sowie die Vektoren um eine 4. Zeile. Die einzelnen Transformationsschritte (Rotation, Translation, Skalierung, Projektion, usw.) werden nacheinander in einer Transformationsmatrix auf multipliziert und gespeichert. Soll ein Punkt nun entsprechend der einzelnen Transformationsschritte transformiert werden, so wird er mit der Matrix multipliziert. Nach der Multiplikation erhält man die Koordinaten des Punktes im 3D-Raum, indem man alle Koordinaten durch die letzte Koordinate teilt. Die Transforma-

tion wird dann in einer Matrix folgendermaßen ausgedrückt:

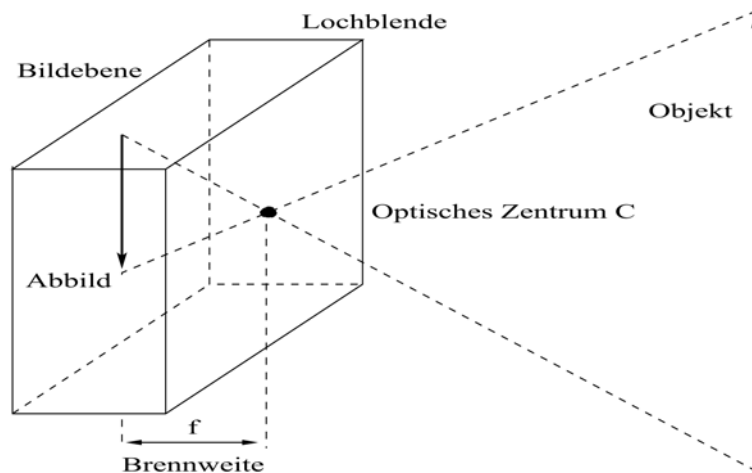
$$\vec{x}' = A\vec{x} = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_1 \\ r_{2,1} & r_{2,2} & r_{2,3} & t_2 \\ r_{3,1} & r_{3,2} & r_{3,3} & t_3 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} x' \\ y' \\ z' \\ 1 \end{pmatrix}$$

Was vorher durch zwei hintereinander ausgeführte Operationen ausgedrückt wurde, kann nun durch eine Matrixmultiplikation mit der Transformationsmatrix  $A$  ausgedrückt werden.  $A$  beinhaltet dabei die Parameter der Rotation, der Translation und der Skalierung. Die Rotation und Skalierung wird durch die Koeffizienten  $r_{1,1}$  bis  $r_{3,3}$  ausgedrückt. Die Translation wird durch die Werte  $t_1, t_2$  und  $t_3$  beschrieben. Wobei die Werte  $t_1, t_2$  und  $t_3$  die Verschiebung in  $x$ -,  $y$ - und  $z$ -Richtung beschreiben. Die vierte Zeile der Matrix zeigt die zuvor beschriebene homogene Erweiterung [15].

## 2.3 Lochkameramodell

Um die reale Welt auf ein zweidimensionales Bild abzubilden, benötigt man ein Modell, mit dem die dreidimensionalen Informationen auf eine zweidimensionale Ebene abgebildet werden. Das einfachste Modell ist dabei das Lochkameramodell, welches alle Sehstrahlen der dreidimensionalen Szene durch ein unendlich kleines Loch (Brennpunkt) bündelt und dort spiegelt. Hinter dem Brennpunkt befindet sich die Bildebene bzw. der Bildsensor, welcher eine zweidimensionale Abbildung der dreidimensionalen Welt aufzeichnet. Ein Sehstrahl wird dabei als Gerade vom Szenenpunkt durch das optische Zentrum der Kamera (Kamerakoordinatensystemursprung) bis zur Bildebene charakterisiert. Das Modell der Lochkamera arbeitet ohne optische Linse und enthält so auch keine optischen Verzerrungen.

Die heute gängigen Aufnahmesysteme spiegeln jedoch kein exaktes Lochkameramodell wieder, sondern besitzen radiale und tangential Abweichungen, welche durch die optischen Linsen im Aufnahmesystem hervorgerufen werden. Um die Verzerrungen zu entfernen, muss man die Ka-



**Abbildung 2.2:** Das Modell der Lochkamera: Einfallendes Licht erzeugt auf der Bildebene ein skaliertes, spiegelverkehrtes Abbild der 3D-Szene vor der Kamera

mera kalibrieren. Dabei werden die kameraspezifischen Parameter, wie z.B. Brennweite und Verzerrungsparameter, in der sogenannten Kameramatrix zusammengefasst. Anschließend berechnet man für jedes einzelne Pixel des Ursprungsbildes die entzerrten Pixelwerte mithilfe der Kameramatrix. Das Normalisieren des Eingangsbildes korrigiert die Abweichungen, sodass das Lochkameramodell wieder verwendet werden kann [5].

## 2.4 Projektion

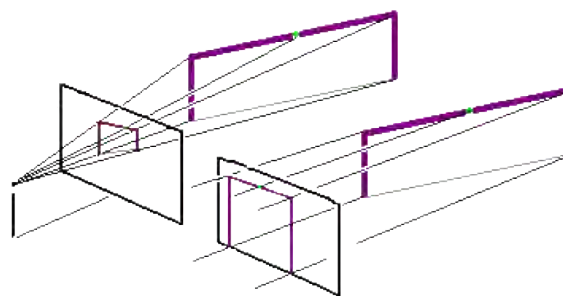
Projektionen transformieren reale 3D-Objekte bzw. Punkte auf eine 2D-Fläche. Allgemein bezeichnet die Projektion eine Transformation von einem höherdimensionalen Raum in einen kleinerdimensionalen Raum. Die Abbildung ist nicht bijektiv, da bei der Projektion die Informationen einer Dimension verloren gehen. In unserem Fall gehen die Tiefeninformationen der 3D-Objekte verloren und können anschließend aus dem Abbild nicht wieder hergestellt werden. Prinzipiell unterscheidet man zwei Arten der Projektion:

**Perspektivisch** Bei der perspektivischen Projektion schneiden sich die Projektionsstrahlen im Projektionszentrum. Die perspektivische bzw. zentrale Projektion arbeitet wie das menschliche Auge und erzeugt so einen natürlichen Bildeindruck. Das zu projizierende Objekt wird



durch die Projektionsstrahlen, welche durch das Projektionszentrum gehen, auf die Bildebene projiziert. Je nach Entfernung zwischen Projektionszentrum und Bildebene wird die Abbildung vergrößert oder verkleinert dargestellt. Die perspektivische Projektion führt zu einer Verzerrung, bei der weiter entfernte Objekte kleiner als näher zum Projektionszentrum platzierte gleich große Objekte abgebildet werden.

**Parallel** Anders als bei der perspektivischen Projektion schneiden sich die Projektionsstrahlen der parallelen Projektion nicht. Die Projektionsstrahlen verlaufen parallel zu einander. Man kann die Parallelprojektion auch als Grenzfall der Zentralprojektion sehen, bei dem der Projektionsursprung im Unendlichen liegt. Im Gegensatz zur perspektivischen Projektion beeinflusst die Entfernung zur Kamera nicht, wie groß ein Objekt dargestellt wird. Gleich große Objekte in der Realität erscheinen auch im projizierten Bild gleich groß. Die parallele Projektion hält die Größe der Objekte unabhängig von der Entfernung zur Kamera bei. Ein Spezialfall der Parallelprojektion ist die senkrechte oder auch orthogonale Projektion. Bei ihr liegen die Projektionsstrahlen ebenfalls parallel zueinander. Zusätzlich verlaufen die Strahlen senkrecht zur Projektionsfläche. Die Projektionstrahlen vom Objekt zur Projektionsfläche haben stets die kürzeste Entfernung zur Projektionsfläche.



**Abbildung 2.3:** Vergleich von Zentralprojektion (links) und Parallelprojektion (rechts) am Beispiel der Projektion eines Torrahmens

Abbildung 2.3 auf der vorherigen Seite zeigt die Unterschiede zwischen der perspektivischen und der parallelen Projektion. Bei der perspektivischen Projektion im linken Bereich des Bildes schneiden sich alle Projektionsstrahlen in einem Punkt, dem Projektionsursprung. Dadurch, dass die Strahlen perspektivisch zusammen laufen, wird der vordere Pfosten, welcher näher zur Bildebene steht, größer abgebildet als der hintere Pfosten, obwohl beide Pfosten in der realen 3D-Welt die selbe Größe haben. Die perspektivische Verzerrung macht sich auch am markierten Mittelpunkt auf der oberen Verbindungsplatte des Tores bemerkbar. Der Punkt ist im Abbild nicht in der Mitte, sondern näher zum hinteren Pfosten.

Der rechte Bereich des Bildes zeigt ein Beispiel für die parallele Projektion. Charakteristisch für die Parallelprojektion verlaufen die Projektionsstrahlen hier parallel zu einander. Dies bedeutet, dass es keine perspektivische Verjüngung weiter entfernter Objekte gibt. Im Bild werden beide Pfosten gleich groß abgebildet. Außerdem wird der Mittelpunkt auf der oberen Latte genau in der Mitte der projizierten Latte abgebildet [10, 11].

### 2.4.1 Mathematische Beschreibung der projektiven Abbildung

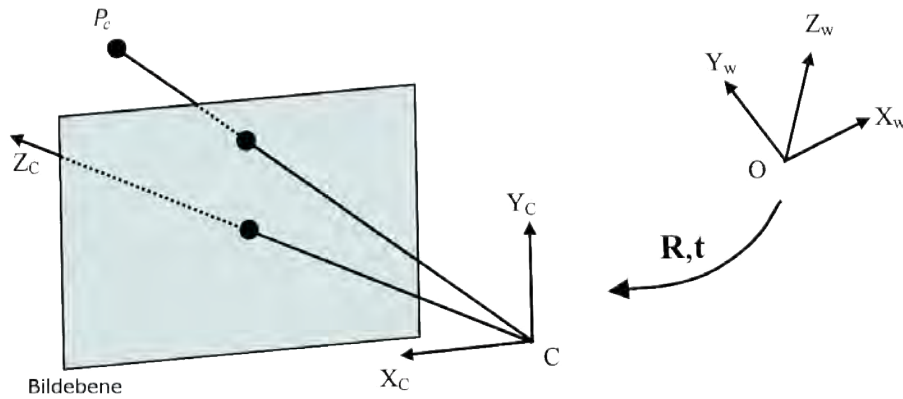
Die projektive Abbildung wird in drei Schritte unterteilt, und zwar in eine externe, eine perspektivische und eine interne Transformation.

#### Externe Transformation

Abbildung 2.4 auf der nächsten Seite zeigt den Schritt der externen Transformation. Ziel ist es, das Weltkoordinatensystem in das Kamerakoordinatensystem zu überführen. Dazu wird eine Transformation bestehend aus Rotation und Translation benötigt.

Die Überführung eines 3D-Weltkoordinatenpunktes  $P_w$  in einen 3D-Kamerakoordinatenpunkt  $P_c$  erfolgt schließlich mit der folgenden Transformation:

$$P_c = AP_w$$



**Abbildung 2.4:** Externe Transformation: Das Weltkoordinatensystem wird in das Kamerakoordinatensystem überführt (aus [10, Seite 42])

Wobei die Transformationsmatrix

$$A = \begin{pmatrix} R & \vec{t} \\ 0 & 1 \end{pmatrix}$$

aus der Rotationsmatrix

$$R = \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} \\ r_{2,1} & r_{2,2} & r_{2,3} \\ r_{3,1} & r_{3,2} & r_{3,3} \end{pmatrix}$$

und dem Translationsvektor

$$\vec{t} = \begin{pmatrix} t_1 \\ t_2 \\ t_3 \end{pmatrix}$$

besteht.

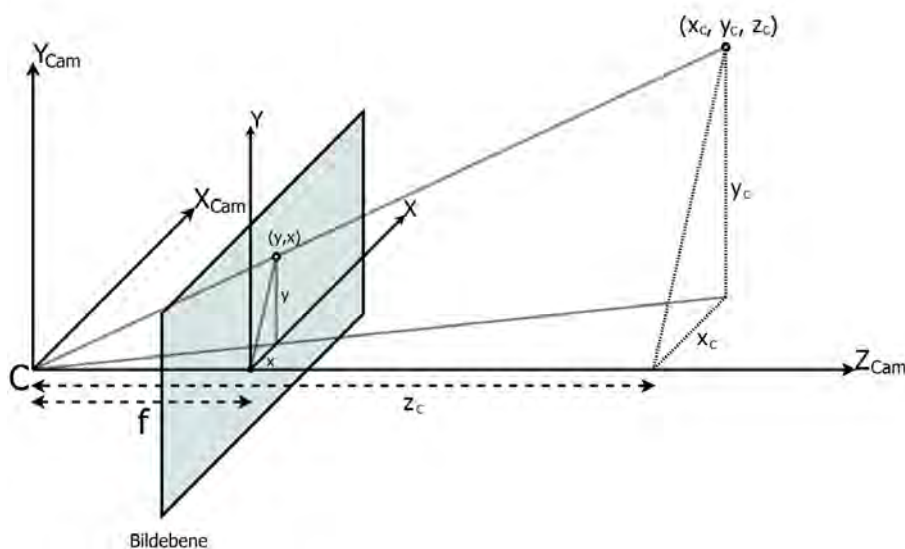
Die Matrix  $\begin{pmatrix} R & \vec{t} \\ 0 & 1 \end{pmatrix}$  wird auch extrinsische Matrix genannt und beschreibt die äußere Orientierung der Kamera durch die Rotation und die Translation zwischen Kamera- und Weltkoordinatensystem.

## Perspektivische Transformation

Die perspektivische Transformation beschreibt die Projektion eines 3D-Punktes im Kamerakoordinatensystem auf die 2D-Bildebene. Die Gleichung der Zentralprojektion legt dabei die Projektionsmatrix durch folgenden Zusammenhang fest:

$$\frac{x}{x_c} = \frac{y}{y_c} = \frac{f}{z_c}$$

Der Zusammenhang zwischen den Bildkoordinaten  $x$  und  $y$  und deren 3D-Koordinaten  $x_c$  und  $y_c$  ergibt sich dabei aus dem 2. Strahlensatz. Dieser besagt: Die Abschnitte auf den Parallelen ( $x$  und  $x_c$  bzw.  $y$  und  $y_c$ ) verhalten sich zueinander wie die Strecken vom Strahlenursprung entlang eines Strahl zu den Parallelen ( $f$  und  $z_c$ ). Abbildung 2.5 macht deutlich, dass bei der perspektivischen Projektion genau dieses Szenario des 2. Strahlensatzes vorliegt [12].



**Abbildung 2.5:** Abbildung eines 3D-Punktes auf die Bildebene durch die perspektivische Projektion

Die perspektivische Transformation wird somit durch folgende Abbil-

ungsgleichung angegeben:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix}$$

Die x- und y-Koordinaten des Bildpunktes erhält man schließlich durch:

$$x' = \frac{x}{z} \quad \text{und} \quad y' = \frac{y}{z} \quad z \neq 0$$

Die Transformation eines 3D-Weltkoordinatenpunktes  $P_w$  in 2D-Bildkoordinaten ergibt sich somit durch Anwendung der externen und der perspektivischen Transformation.

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} R & \vec{t} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}$$

### Interne Transformation

Mithilfe der internen Transformation werden nun die metrischen Koordinaten in diskrete Bildkoordinaten umgerechnet. Dies geschieht mithilfe einer Skalierung entlang der horizontalen ( $k_u$ ) und der vertikalen ( $k_v$ ) Achsen. Neben der Skalierung wird in der Regel noch eine Verschiebung des Schnittpunktes ( $u_0, v_0$ ) der optischen Achse mit der Bildebene (Kamerahauptpunkt) durchgeführt. Dies ist üblich, da der Ursprung der Bildmatrix üblicherweise in der oberen linken Ecke des Bildes liegt. Es ergibt sich folgende homogene Transformationsmatrix:

$$H = \begin{pmatrix} k_u & 0 & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

In Kombination mit der perspektivischen Transformation ergibt sich folgende Matrix:

$$K = \begin{pmatrix} fk_u & 0 & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

Allgemein wird die Abbildung vom 3D-Weltkoordinatenpunkt auf die Pixelmatrix beschrieben durch:  $P_b = KAP_w$ . Wobei  $P_b$  ein Bildpunkt der Pixelmatrix,  $P_w$  ein 3D-Punkt im Weltkoordinatensystem,  $A$  die Transformationsmatrix vom 3D-Welt- ins 3D-Kamerakoordinatensystem und  $K$  die Abbildungsmatrix der 3D-Kamerapunkte auf die Bildebene beinhaltet:

$$P_b = KAP_w = \begin{pmatrix} fk_u & 0 & u_0 \\ 0 & fk_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R & \vec{t} \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}$$

[16, 10]

## 2.5 Kameraparameter

Die Kameraparameter unterscheiden sich von Kameramodell zu Kameramodell. Sie lassen sich einteilen in Parameter der äußeren und inneren Orientierung. Dabei beschreibt die äußere Orientierung die Position und die Ausrichtung der Kamera im Raum. Die Parameter der äußeren Orientierung werden als extrinsische Parameter bezeichnet. In der Regel wird die Orientierung in einer Rotationsmatrix und einem Translationsvektor ausgedrückt (Vergleiche Kapitel 2.4.1 auf Seite 13). Mit dem Thema äußere Orientierung befasst sich speziell das Kapitel 3 auf Seite 21.

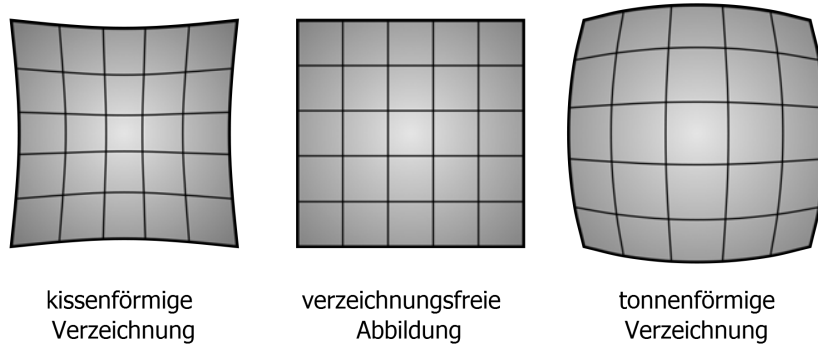
Die innere Orientierung beinhaltet die kameraspezifischen Größen, welche unter anderem zur Entzerrung des Bildes benötigt werden. Neben der Brennweite  $f$  sind dies der Bildhauptpunkt  $c$  und die je nach Modell unterschiedlichen Verzerrungskoeffizienten. Die Parameter der inneren Orientierung werden als intrinsische Parameter der Kamera bezeichnet und definieren die Abbildung zwischen dem 3D-Kamera- und dem 2D-Bildkoordinatensystem. Man erhält mit ihnen Informationen über die interne Geometrie der Kamera. Man benötigt die Parameter, um die Verzerrung durch die Linse rückgängig zu machen. Die Linsenverzerrung beinhaltet dabei zwei Arten von geometrischer Verzerrung:

**Radiale Verzerrung** Die radiale Verzerrung wird durch die Kameralinse verursacht. Die radiale Verzerrung hängt in ihrer Intensität von der Entfernung zum Linsenmittelpunkt ab. In der Mitte der Linse ist die Linse nicht gewölbt und die Strahlen können unverändert wieder austreten. Je weiter man sich jedoch vom Mittelpunkt entfernt, desto stärker machen sich Linsenfehler bemerkbar. Die radiale Verzerrung hat den Effekt, dass Geraden nicht mehr auf Geraden sondern gekrümmt (kissenförmig oder tonnenförmig) abgebildet werden (Vergleiche Abbildung 2.6). Zur Rückrechnung der radialen Verzerrung benötigt man 3 Verzerrungskoeffizienten ( $k_1, k_2, k_3$ ). Die korrigierte Pixel-Position für das Pixel  $(x, y)$  erhält man durch folgende Polynom Approximations-Formel:

$$x_{new} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)$$

$$y_{new} = y(1 + k_1r^2 + k_2r^4 + k_3r^6)$$

mit  $r$  als Radius bzw. Abstand zum Mittelpunkt ( $r^2 = x^2 + y^2$ ) [21, 6].



**Abbildung 2.6:** Auswirkungen der radialen Verzerrung(aus [16]).

**Tangentiale Verzerrung** Verläuft die optische Achse nicht senkrecht durch die Kameralinse, so kommt es zur tangentialen Verzerrung. Dies hat den Effekt, dass die Punkte entlang der Tangente an den Kreis durch den Punkt um den Kamerahauptpunkt verschoben werden. Die Verzerrung wird mit der Entfernung zum Kamerahauptpunkt größer. Die tangentielle Verzerrung hat jedoch im Vergleich zur radialen Verzerrung einen geringeren Einfluss. Die Rückrechnung der tangentialen Verzerrung benötigt 2 Verzerrungskoeffizienten ( $p_1, p_2$ ). Die Formel für die korrigierten Pixel-Koordinaten lautet:

$$x_{new} = x + [2p_1y + p_2(r^2 + 2x^2)]$$

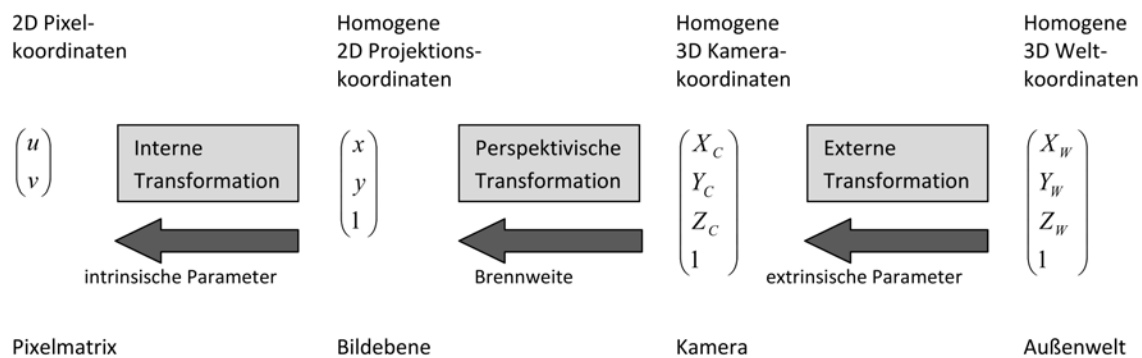
$$y_{new} = y + [2p_1(r^2 + 2y^2) + 2p_2x]$$

[21]

Die intrinsischen Parameter der Kamera enthält man durch die Kalibration der Kamera mit Kalibrationsmustern. Dabei werden bekannte Prüfmuster aufgenommen und anschließend anhand der Abweichung zum Original die Verzerrungskoeffizienten berechnet (vergleiche Kapitel 3 auf Seite 21).



Abbildung 2.7 zeigt zusammenfassend an welcher Stelle in der Transformationspipeline die Kameraparameter einfließen. Die extrinsischen Parameter werden zur externen Transformation der 3D-Weltkoordinaten ins 3D-Kamerakoordinatensystem benötigt. Zur perspektivischen Transformation der 3D-Kamerakoordinaten in 2D-Bildkoordinaten benötigt man die Brennweite der Kamera. Anschließend werden die intrinsischen Parameter der Kamera zur internen Transformation der homogenen 2D-Bildkoordinaten in Pixelkoordinaten benötigt. [10, 18].



**Abbildung 2.7:** Die Abbildung zeigt die Transformation zwischen 3D-Weltkoordinaten und 2D-Pixelbild.

# 3 Kamera Kalibrierung

Die Kalibrierung der Kamera besteht daraus, die inneren und äußeren Kameraparameter des verwendeten Kameramodells zu bestimmen. Dazu werden mehrere Punktpaare aus der 2D-Bildebene und dem 3D-Weltkoordinatensystem benötigt. Wie bereits in Kapitel 2.5 auf Seite 18 beschrieben, werden die Kameraparameter in Parameter der inneren (intrinsischen) und äußeren (extrinsischen) Orientierung unterteilt. Kalibrierung bezeichnet dabei den Prozess der Bestimmung der nötigen Modellparameter für eine bestimmte Kamera.

Die extrinsischen Parameter beschreiben die äußere Orientierung der Kamera, welche die Position und die Ausrichtung der Kamera bezüglich des Weltkoordinatensystems beinhaltet. Das optische Zentrum entspricht dabei der Position der Kamera. Die z-Achse des Kamerakoordinatensystems entspricht der optischen Achse bzw. der Richtung in die die Kamera ausgerichtet ist.

Die intrinsischen Parameter hingegen beschreiben die interne Geometrie der Kamera. Sie definieren die Abbildung zwischen dem metrischen 3D-Kamerakoordinatensystem und dem 2D-Bildkoordinatensystem. Sie sorgen für die Projektion der 3D-Punkte auf die 2D-Bildebene und die anschließende Diskretisierung in Bildpixel. Die zu bestimmenden intrinsischen Parameter sind die Brennweite  $f$  (Entfernung vom Zentrum der Projektion zur Bildebene), die Pixelkoordinaten der Bildmitte  $(u_0, v_0)$  sowie die Pixelskalierung (das Verhältnis von Pixel zu Meter) in x- und y-Richtung  $(k_u, k_v)$ . Zusammengefasst ergeben die intrinsischen Parameter die Kameramatrix:

$$K = \begin{pmatrix} f k_u & 0 & u_0 \\ 0 & f k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

Das Bestimmen der Parameter der inneren Orientierung wird mithilfe von Kalibrationsmustern durchgeführt. Dazu werden mehrere Aufnahmen eines Kalibrationsmusters mit der zu kalibrierenden Kamera aufgenommen. Das Kalibrationsmuster kann dabei aus einem Schachbrettmuster (vergleiche Abbildung 3.1) bestehen. Die Kalibrationsroutine der Program-



**Abbildung 3.1:** Schachbrettmuster zur Kamerakalibrierung (links), Grafische Ausgabe der Kalibrationsfunktion (rechts)

mierbibliothek OpenCV erhält dazu Informationen, wie viele Felder das Schachbrett besitzt. Mit der Kenntnis, dass die Felder quadratisch und exakt gerade angeordnet sind, berechnen die Routinen die internen Parameter wie Brennweite, Bildhauptpunkt und die Parameter der radialen und tangentialen Verzerrung. Für die Anwendung des Programms wird im weiteren Verlauf davon ausgegangen, dass man mit entzerrten Bildern arbeitet. Dazu muss das Bild, wie in Kapitel 2.5 auf Seite 18 beschrieben, entzerrt werden.

Nachdem die internen Parameter durch die OpenCV Kalibrationsroutinen bestimmt wurden, beginnt das Verfahren der Kalibration der äußeren Kameraparameter. Mithilfe der Punktkorrespondenzen zwischen 3D-Welt und 2D-Bildkoordinaten sowie den intrinsischen Kameraparametern wird im folgenden gezeigt, wie man die extrinsischen Kameraparameter berechnen kann [18, 21].

## 3.1 Perspective-n-Point Problem (PnP-Problem)

Das Problem, welches sich mit diesem Thema beschäftigt, ist das Perspective-n-Point Problem (PnP-Problem). Ziel des Perspective-n-Point Problems ist es, aus  $n$  gegebenen Punktkorrespondenzen die Kameraposition und -ausrichtung zur 3D-Szene zu ermitteln.

Es gibt verschiedene Ansätze und Verfahren das PnP-Problem zu lösen. Exemplarisch soll im Folgenden anhand des POSIT-Verfahrens eine Vorstellung vermittelt werden, wie man die extrinsischen Parameter aus den Punktkorrespondenzen bestimmen kann. Die Bestimmung der extrinsischen Kameraparameter spielt eine wichtige Rolle, da viele Anwendungen im Bereich der Kartographie, Tracking und Objekterkennung darauf aufbauen [17].

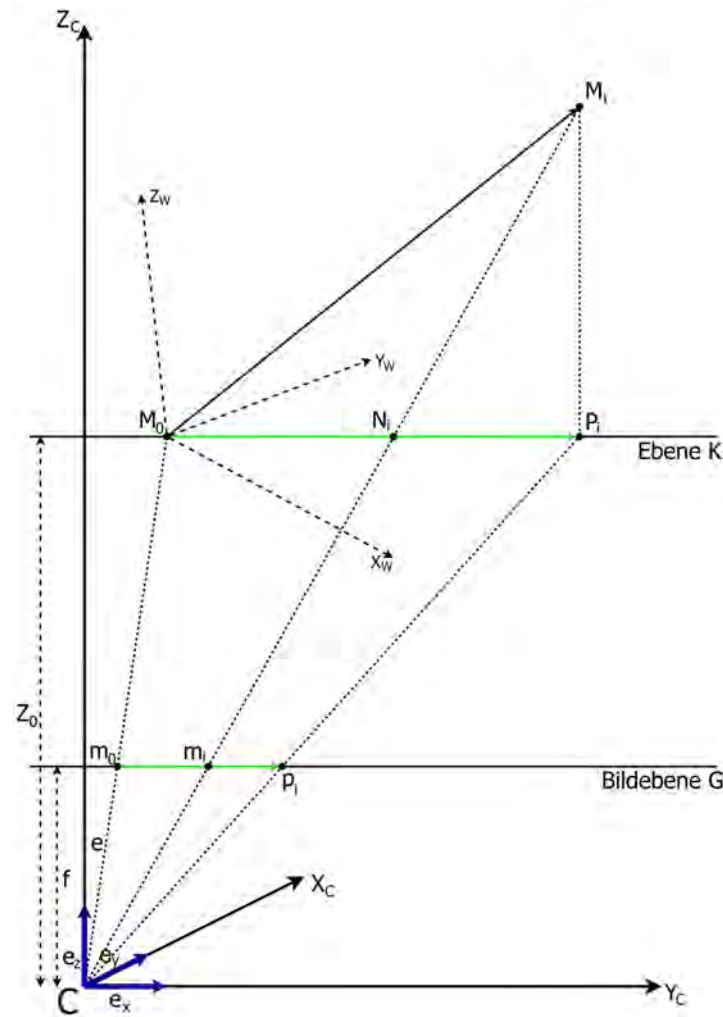
### 3.1.1 POSIT

Der POSIT-Algorithmus bietet eine Möglichkeit, die räumliche Lage eines Objektes aus einem einzelnen aufgenommenen Bild im Koordinatensystem der Kamera zu bestimmen. Dazu benötigt man mindestens vier nicht in einer Ebene liegende Punkte des Objektes. Die relative Lage der Punkte zueinander muss bekannt sein. Neben den 3D-Weltkoordinaten benötigt man ebenfalls die 2D-Bildkoordinaten, welche die 3D-Punkte auf der Bildebene abbilden. Zudem wird die Brennweite der Projektion benötigt, die durch die vorherige Kalibrierung bestimmt wurde oder so eingestellt werden muss, dass eine Hin- und Rückprojektion der Punkte der Punktkorrespondenzen möglichst wenig abweicht.

#### Vorbemerkungen

Gesucht sind die Rotationsmatrix  $R$  und der Translationsvektor  $\vec{t}$ . Der Ursprung des Kamerakoordinatensystems wird mit  $C$  bezeichnet und die Achsen mit  $X_C$ ,  $Y_C$  und  $Z_C$  mit den entsprechenden Einheitsvektoren  $\vec{e}_x$ ,  $\vec{e}_y$  und  $\vec{e}_z$ . Die im Sichtfeld der Kamera liegenden 3D-Punkte werden

mit  $M_0$  bis  $M_i$ ,  $i \in \mathbb{N}_n$  bezeichnet, wobei  $n$  die Anzahl der Punktkorrespondenzen angibt. Der Punkt  $M_0$  ist zugleich der Koordinatenursprung des Weltkoordinatensystems. Die perspektivische Projektion des Punktes  $M_i$  wird auf der Bildebene  $G$  mit  $m_i$  bezeichnet. Die skalierte senkrechte Projektion des Punktes  $M_i$  wird auf der Bildebene mit  $p_i$  dargestellt (vergleiche Abbildung 3.2).



**Abbildung 3.2:** Vergleich der perspektivischen Projektion ( $m_i$ ) und der skalierten senkrechten Projektion ( $p_i$ ) des Punktes  $M_i$  auf die Bildebene  $G$

Die Rotationsmatrix und der Translationsvektor überführen einen Punkt aus dem Weltkoordinatensystem in Koordinaten des Kamerakoordinatensystems. Die Zeilen der Rotationsmatrix beinhalten die Koordinaten der Einheitsvektoren des Kamerakoordinatensystems, ausgedrückt im Welt-

koordinatensystem. Deutlich wird dies, wenn man die Skalarmultiplikation des Vektors  $\overrightarrow{M_0M_i}$ , welcher von  $M_0$  zu  $M_i$  zeigt, mit dem Einheitsvektor  $\vec{e}_x$  betrachtet. Das Resultat liefert den Anteil des Vektors  $\overrightarrow{M_0M_i}$  auf der  $x$ -Achse des Kamerakoordinatensystems. Genau der Wert, welcher der  $x$ -Koordinate im Kamerakoordinatensystem entspricht. Die Rotationsmatrix  $R$  enthält somit die Einheitsvektoren des Kamerakoordinatensystems ausgedrückt in Weltkoordinaten.

$$R = \begin{pmatrix} \vec{e}_x^T \\ \vec{e}_y^T \\ \vec{e}_z^T \end{pmatrix}$$

Der Vektor  $\vec{e}_z$  berechnet sich aus dem Kreuzprodukt von  $\vec{e}_x$  und  $\vec{e}_y$ , da er orthogonal zu den beiden Vektoren steht:

$$\vec{e}_z = \vec{e}_x \times \vec{e}_y$$

Der Translationsvektor  $\vec{t}$  wird durch den Verbindungsvektor zwischen dem Projektionsursprung  $C$  und dem Punkt  $M_0$  ausgedrückt. Da der Translationsvektor die selbe Richtung wie der Vektor  $\overrightarrow{Cm_0}$  hat, kann der Translationsvektor folgendermaßen angegeben werden:

$$\vec{t} = \frac{1}{s} \overrightarrow{Cm_0}$$

mit dem Skalierungsfaktor  $s = \frac{f}{z_{c(0)}}$  der skalierten senkrechten Projektion. Der Translationsvektor kann somit wie folgt berechnet werden:

$$\vec{t} = \frac{z_{c(0)}}{f} \overrightarrow{Cm_0}$$

Da der Projektionsursprung  $C$ , die Koordinaten des Bildpunktes  $m_0$  sowie die Brennweite  $f$  bekannt sind, benötigt man nur  $z_{c(0)}$ , die  $z$ -Komponente des Punktes  $M_0$ .

Die Parameter der externen Transformation sind vollständig bestimmbar, sobald man die Werte der Vektoren von  $\vec{e}_x$ ,  $\vec{e}_y$  und den Wert  $z_{c(0)}$  des Punktes  $M_0$  berechnet hat.

### Skalierte Senkrechte Projektion (SOP)

Die skalierte senkrechte Projektion (SOP = scaled orthographic projection) beschreibt ein Verfahren zur Annäherung der perspektivischen Projektion. Dabei wird davon ausgegangen, dass die  $z$ -Koordinaten (Tiefe) der verwendeten Kalibrationspunkte im Verhältnis zur Entfernung zur Kamera, nur gering voneinander abweichen.

Zunächst wird jeder Punkt senkrecht auf eine Ebene  $K$  projiziert, welche parallel zur Bildebene durch den Punkt  $M_0$  geht (vergleiche Abbildung 3.2 auf Seite 24). Dies hat zur Folge, dass alle 3D-Punkte nun die selbe  $z$ -Koordinate  $z_{c(0)}$  haben. Die so erhaltenen Punkte  $P_i$  auf der Ebene  $K$  werden anschließend durch eine perspektivische Projektion auf die Bildebene projiziert. Die Projektion des Punktes  $M_i = \begin{pmatrix} x_{c(i)} \\ y_{c(i)} \\ z_{c(i)} \end{pmatrix}$  wird bei der perspektivischen Projektion durch folgende Formel ausgedrückt:

$$x_i = f \frac{x_{c(i)}}{z_{c(i)}}, \quad y_i = f \frac{y_{c(i)}}{z_{c(i)}},$$

bei Anwendung der skalierten senkrechten Projektion ergibt sich folgende Projektionsformel:

$$x'_i = f \frac{x_{c(i)}}{z_{c(0)}}, \quad y'_i = f \frac{y_{c(i)}}{z_{c(0)}},$$

wobei  $m_i = \begin{pmatrix} x_i \\ y_i \end{pmatrix}$  den Bildpunkt der perspektivischen Projektion und  $p_i = \begin{pmatrix} x'_i \\ y'_i \end{pmatrix}$  den Bildpunkt der skalierten senkrechten Projektion wiedergibt. Das Verhältnis  $s = \frac{f}{z_{c(0)}}$  wird als Skalierung, gemäß Strahlensatz, der skalierten senkrechten Projektion bezeichnet und beschreibt das Skalierungsverhältnis zwischen dem Vektor  $\overrightarrow{M_0 P_i}$  auf der Ebene  $K$  und dem Vektor  $\overrightarrow{m_0 p_i}$  auf der Bildebene  $G$ .

### Vorgehensweise

Das POSIT-Verfahren wird in zwei Schritte unterteilt. Zunächst wird durch Approximation der Perspektivischen Projektion durch die skalier-

te senkrechte Projektion ein Startwert für die Rotationsmatrix und den Translationsvektor bestimmt. Mit diesen Angaben der äußeren Orientierung wird eine erste Näherung der Positionen der Punkte berechnet. Dieser Schritt des Verfahrens wird POS genannt (Pose from Orthography and Scaling). Im zweiten Schritt wird die berechnete äußere Orientierung nun verwendet, um bessere Werte für die skalierte senkrechte Projektion der Objektpunkte zu erhalten. Die Werte werden iterativ verbessert bis sie sich unter einer angegebenen Genauigkeit nicht mehr ändern. Das gesamte Verfahren nennt man schließlich POSIT (POS with ITerations) [7, 14].

### Mathematischer Hintergrund

Wir betrachten zunächst folgende Gleichungen:

$$\overrightarrow{M_0 M_i} \vec{E}_x = x_i(1 + \epsilon_i) - x_0,$$

$$\overrightarrow{M_0 M_i} \vec{E}_y = y_i(1 + \epsilon_i) - y_0,$$

mit

$$\vec{E}_x = \frac{f}{z_{c(0)}} \vec{e}_x, \quad \vec{E}_y = \frac{f}{z_{c(0)}} \vec{e}_y, \quad \epsilon_i = \frac{1}{z_{c(0)}} \overrightarrow{M_0 M_i} \vec{e}_z$$

und  $\vec{e}_z = \vec{e}_x \times \vec{e}_y$ . Die linke Seite der Gleichung beschreibt dabei die skalierte senkrechte Projektion des Vektors  $\overrightarrow{M_0 M_i}$  auf den Einheitsvektor  $\vec{e}_x$  bzw.  $\vec{e}_y$ . Auf der rechten Seite ist entsprechend der  $x$ - bzw.  $y$ -Anteil des projizierten Vektors auf der Bildebene dargestellt. Der Ausdruck  $x_i(1 + \epsilon_i)$  bzw.  $y_i(1 + \epsilon_i)$  entspricht dabei den  $x'_i$ - und  $y'_i$ -Koordinaten der skalierten senkrechten Projektion des Punktes  $M_i$ . Der Wert für  $\epsilon_i$  wird zunächst auf 0 gesetzt. Was bedeutet, dass die Bildpunkte  $m_i$  und  $p_i$  für die perspektivische und die skalierte senkrechte Projektion übereinstimmen.

Man erhält ein Gleichungssystem, indem nur die Werte für  $\vec{E}_x$  und  $\vec{E}_y$  zu bestimmen sind. Hat man  $\vec{E}_x$  und  $\vec{E}_y$  einmal berechnet, so erhält man die



Einheitsvektoren  $\vec{e}_x$  und  $\vec{e}_y$  durch Normalisieren von  $\vec{E}_x$  und  $\vec{E}_y$ .

$$\vec{e}_x = \frac{\vec{E}_x}{|\vec{E}_x|}, \quad \vec{e}_y = \frac{\vec{E}_y}{|\vec{E}_y|}$$

Die Güte der ersten Approximation der äußeren Orientierung wird durch  $\epsilon_i$  ausgedrückt.  $\epsilon_i$  beschreibt dabei das Verhältnis zwischen der Entfernung der Kalibrationspunkte zur Kamera und den unterschiedlichen Tiefen der berechneten Punkte  $M_i$ . Ist die Entfernung zur Kamera relativ groß und der  $z$ -Abstand der Objektpunkte gering, so wäre der Wert für  $\epsilon_i$  klein und eine erste Näherung würde bereits brauchbare Ergebnisse liefern. Liefert  $\epsilon_i$  einen großen Wert, so wird die projizierte Bildposition  $p_i$  der skalierten senkrechten Projektion entsprechend weit verschoben, bis sich die Position bzw.  $\epsilon_i$  unter einer angegebenen Genauigkeit nicht mehr verändert und man die Bildposition  $p_i$  der skalierten senkrechten Projektion für die Punkte  $M_i$  gefunden hat.

In verkürzter Schreibweise erhält man folgende Gleichungssysteme:

$$Q\vec{E}_x = \vec{x}', \quad Q\vec{E}_y = \vec{y}'$$

wobei  $Q$  eine  $n \times 3$ -Matrix ist und zeilenweise die Vektoren  $\overrightarrow{M_0M_i}$  vom Referenzpunkt  $M_0$  zu den Objektpunkten  $M_i$  enthält. Die Vektoren  $\vec{x}'$  und  $\vec{y}'$  enthalten die Koordinaten der Punkte  $p_i$  für die skalierte senkrechte Projektion. Zur Lösung des Gleichungssystem wird die Pseudoinverse  $B$ , welche im folgenden auch Objektmatrix genannt wird, verwendet. Sie wird vor der Iteration einmalig durch folgende Operation gebildet:

$$B = [Q^T Q]^{-1} Q^T$$

Es folgt:

$$\vec{E}_x = B\vec{x}', \quad \vec{E}_y = B\vec{y}'$$

Alternativ zu der Objektmatrix kann man auch die Singulärwertzerlegung zur Lösung des Gleichungssystems verwenden [24, 25].

Mithilfe dieser Berechnungen erhält man die Vektoren  $\vec{E}_x$  und  $\vec{E}_y$ . Die

Einheitsvektoren  $\vec{e}_x$  und  $\vec{e}_y$  fallen durch Normalisieren der Vektoren ab. Den Skalierungsfaktor  $s$  der skalierten senkrechten Projektion erhält man, indem man den Mittelwert der beiden Vektorlängen bildet:

$$s = \frac{|\vec{e}_x| + |\vec{e}_y|}{2}$$

Da  $\vec{e}_z$  durch das Kreuzprodukt von  $\vec{e}_x$  und  $\vec{e}_y$  berechnet werden kann, hat man alle Elemente der Rotationsmatrix berechnet. Der Translationsvektor  $\vec{t} = \overrightarrow{CM_0}$  wird durch  $\vec{t} = \frac{z_{c(0)}}{f} \overrightarrow{Om_0}$  definiert.  $z_{c(0)}$  erhält man schließlich durch das Skalierungsverhältnis  $s = \frac{f}{z_{c(0)}} \Leftrightarrow z_{c(0)} = \frac{f}{s}$  [7, 14, 8].



## 4 Implementierung

### 4.1 Verfahrensbeschreibung

Folgend eine Übersicht über die einzelnen Schritte, welche von der Aufnahme des Bildes bis zur Dichtebestimmung durchlaufen werden.

#### Voraussetzungen

Bei der Aufnahme des zu untersuchenden Bildes ist es hilfreich die Parameter der Kameramatrix wie z.B. die Brennweite mit der unveränderten Kamera zu bestimmen. Dazu muss die Kamera anhand eines Kalibrationsmusters kalibriert werden. Mithilfe der Kameraparameter wird das Bild anschließend entzerrt. Das Verfahren geht später von einem bereits entzerrten Bild aus. Ist eine Kalibration der Kamera nicht möglich, so muss die Brennweite experimentell durch Ausprobieren verschiedener Brennweiten möglichst optimal angenähert werden. Die weiteren Kameraparameter werden in diesem Fall nicht berücksichtigt und auf Standardwerte gesetzt.

#### Kalibration

Die Kalibration in diesem Verfahren besteht daraus, die äußere Orientierung der Kamera zu bestimmen. Dazu müssen einige Punkte im 2D-Bild und die dazugehörigen 3D-Punkte in der Szene ausgemessen werden. Die Geometrie unter den 3D-Punkten muss bekannt sein. Die Kalibration erfolgt mithilfe der OpenCV-Funktion `solvePnP`, welche die Punktkorrespondenzen und die Elemente der Kameramatrix verwendet, um die Rotation

und die Translation zwischen Kamerakoordinatensystem und Weltkoordinatensystem zu berechnen. Wie die Berechnung der äußeren Orientierung im einzelnen funktioniert wird in Kapitel 3.1 auf Seite 23 genauer beschrieben.

## Projektion

Die berechnete Rotation und Translation ermöglicht es nun, zu einem 3D-Punkt die entsprechende 2D-Projektion auf der Bildebene anzugeben. Neben der äußeren Orientierung benötigt man die innere Orientierung der Kamera um eine Projektion auf die Bildebene durchführen zu können.

## Positionsbestimmung

Für die Positionsbestimmung ist der umgekehrte Weg vom 2D-Bildpunkt zu den 3D-Koordinaten nötig, um die Position im Raum zu bestimmen. Dazu betrachtet man die Gerade vom Projektionsursprung durch den Bildpunkt in die 3D-Szene hinein, auf welcher der gesuchte 3D-Punkt liegt. Wie in der Aufgabenstellung bereits beschrieben wird für die Bestimmung der Kopfpositionen davon ausgegangen, dass sich alle Personen auf einer Ebene befinden und damit auch alle Köpfe idealerweise nahezu in einer Ebene liegen. Der Schnittpunkt der Geraden mit der Ebene auf Höhe der Personenköpfe wird als Position des Kopfes angenommen. Der Schnittpunkt wird auf den Boden projiziert, um den Standpunkt der Person zu erhalten.

## Dichtebestimmung

Die Positionen der Personen auf der Ebene werden anschließend als Ausgangspunkt eines Voronoi-Diagramms genommen. Die einzelnen Voronoi-Zellen beschreiben den jeder Person am nächsten liegenden Bereich und stellen dabei den persönlichen Raum dar. Die Fläche der Voronoi-Zelle ist dabei umgekehrt proportional zur Dichte in dieser Zelle. Das resultierende Voronoidiagramm aller Voronoi-Zellen gibt eine Übersicht über die Dichteverteilung in der Menschenmasse[20].

## 4.2 Kalibrierung

Die Kalibrierung besteht bei der entwickelten Software daraus, die äußere Orientierung der Kamera zur aufgenommenen Szene zu ermitteln. Dazu benötigt man mehrere ausgemessene 3D-Punkte und ihre korrespondierenden 2D-Bildkoordinaten. Desweiteren werden die Parameter der Kameramatrix benötigt.

Die Funktionen der OpenCV-Bibliothek verwenden dabei das in Kapitel 2.2 auf Seite 11 beschriebene Modell der Lochkamera. Die Abbildung durch die Lochkamera wird durch folgende Gleichung ausgedrückt:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & u_0 \\ 0 & f & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} r_{1,1} & r_{1,2} & r_{1,3} & t_1 \\ r_{2,1} & r_{2,2} & r_{2,3} & t_2 \\ r_{3,1} & r_{3,2} & r_{3,3} & t_3 \end{pmatrix} \begin{pmatrix} x_w \\ y_w \\ z_w \\ 1 \end{pmatrix}$$

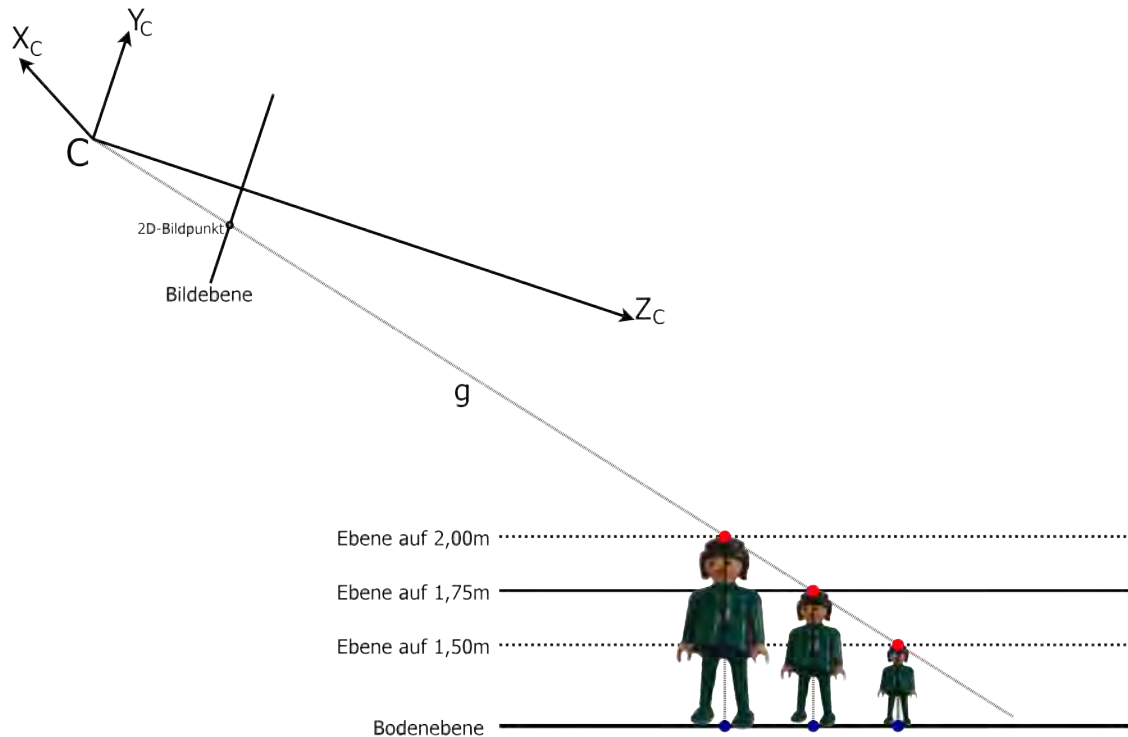
Die OpenCV-Funktion `solvePnP` erwartet als Eingabeparameter die Koordinaten der Punktkorrespondenzen aus 2D-Bild und der 3D-Szene sowie die Kameramatrix. Optional können auch noch Verzerrungskoeffizienten angegeben werden, falls das Bild noch nicht vorab entzerrt wurde. Als Ergebnis erhält man die äußere Orientierung der Kamera in einer Rotationsmatrix  $R$  und einem Translationsvektor  $\vec{t}$ . Die Funktion minimiert dazu den Fehler zwischen den im 2D-Bild gemessenen Bildpunkten und den durch die berechnete Orientierung auf die Bildebene projizierten 3D-Punkten.

## 4.3 Positionsberechnung

Dank der berechneten äußeren Orientierung der Kamera kann man nun Berechnungen im Kamerakoordinatensystem durchführen. Bildlich kann man sich das Vorgehen wie folgt vorstellen.

Nachdem man am Bildschirm den Bildpunkt für den Kopf einer Person ausgewählt hat, legt man eine Gerade vom Projektionsursprung durch den Bildpunkt in die 3D-Szene. Da die Tiefeninformation fehlt, benötigt man

weitere Informationen, um Aufschluss über die Länge der Strecke bis zum 3D-Punkt zu erhalten. Die zusätzliche Information kommt in diesem Fall durch die angenommene Personengröße. Die Gerade geht so weit in die



**Abbildung 4.1:** Schnittpunktberechnung der Geraden vom Projektionsursprung durch den Bildpunkt mit der Ebene auf Höhe der Personenköpfe

Tiefe des Bildes bis sie die Ebene auf Höhe der Personengröße schneidet. Dieser Schnittpunkt gibt die Position des Kopfes der Person wieder. Um nun von der Position des Kopfes der Person auf die Position auf dem Boden zu gelangen, muss noch das senkrechte Lot gefällt werden.

Abbildung 4.1 soll dabei das Vorgehen zur Positionsbestimmung verdeutlichen. Die Gerade  $g$  beginnt am Projektionsursprung  $C$  und geht durch den markierten Punkt auf der Bildebene. Die zu schneidende Ebene liegt parallel zum Boden auf Höhe der angenommenen Personengröße. Der Schnittpunkt der Geraden mit der Ebene (roter Punkt) hängt von der angenommenen Größe (2 m, 1,50 m, 1 m) der Personen ab. Entsprechend wandert die berechnete Standortposition (blauer Punkt) weiter nach vorne (groß) oder weiter zurück (klein).

Die Gerade vom Projektionsursprung durch den Bildpunkt erhält man, indem man zwei Punkte dieser Geraden auswählt und daraus den Richtungsvektor der Geraden berechnet. Als Punkte wählt man den Projektionsursprung sowie einen Punkt im Raum vor der Kamera, welcher auf der Geraden liegt. Diesen bekommt man, indem man den Bildpunkt ins Kamerakoordinatensystem überträgt. Dazu wird die Projektion des Bildpunktes rückgängig gemacht. Die Projektion erfolgte durch folgende Matrix-Multiplikation:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = \begin{pmatrix} fx_c \\ fy_c \\ z_c \end{pmatrix}$$

$$\text{mit } \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{fx_c}{z_c} \\ \frac{fy_c}{z_c} \end{pmatrix}$$

$$\text{damit ergibt sich für } P_c = \begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = \begin{pmatrix} \frac{x}{f} z_c \\ \frac{y}{f} z_c \\ z_c \end{pmatrix}$$

wobei  $z_c$  beliebig gewählt werden kann und den Wert der Tiefe des Punktes entlang der optischen Achse definiert.

Anschließend muss man noch die Ebene definieren, welche die Höhe der Köpfe der Personen angibt. Günstig ist es in dem Fall, wenn das Weltkoordinatensystem so definiert ist, dass die  $XY$ -Ebene des Koordinatensystems die Ebene ist, auf der die Personen sich bewegen. In diesem Fall würde man die  $XY$ -Ebene nur entsprechend der Körpergröße entlang der  $Z_W$ -Achse verschieben. Falls dies nicht der Fall ist, sollte man bei der Vermessung der 3D-Punkte berücksichtigen, dass man hinterher die Bewegungsebene trotzdem angeben kann, indem man z.B. einige (mindestens drei) Punkte der Ebene relativ zum Weltkoordinatensystem mit ausmisst. Die Personengröße, also auf welcher Höhe sich die Ebene zum Schneiden der Geraden befindet, lässt sich im Programm beliebig variieren, ist jedoch für alle Personen im Bild identisch. Größenunterschiede unter den Personen führen so zu fehlerhaften Resultaten, da sie nicht berücksichtigt werden. Eine Untersuchung des Fehlers bzw. des Verhaltens für eine zu groß oder zu klein angenommene Größe wird in Kapitel 5.3 auf Seite 52



diskutiert.

Um die Position der Person auf dem Boden zu erhalten ist es noch nötig, den Schnittpunkt auf die entsprechende Ebene unter der Person senkrecht zu projizieren.

## 4.4 Dichtebestimmung

### Voronoizellenberechnung

Um eine sinnvolle Größe für die Dichte einer einzelnen Person zu ermitteln, wird für jede Person ihre Voronoi-Zelle bestimmt. Die Voronoi-Zelle umfasst dabei den Bereich, welcher der Person am nächsten liegt im Vergleich zu allen weiteren Personen. Die Fläche innerhalb der Voronoi-Zelle einer Person beschreibt den persönlichen Raum dieser Person. Das Bilden der Voronoizellen führt zu einer konvexen Zerlegung des Raumes.

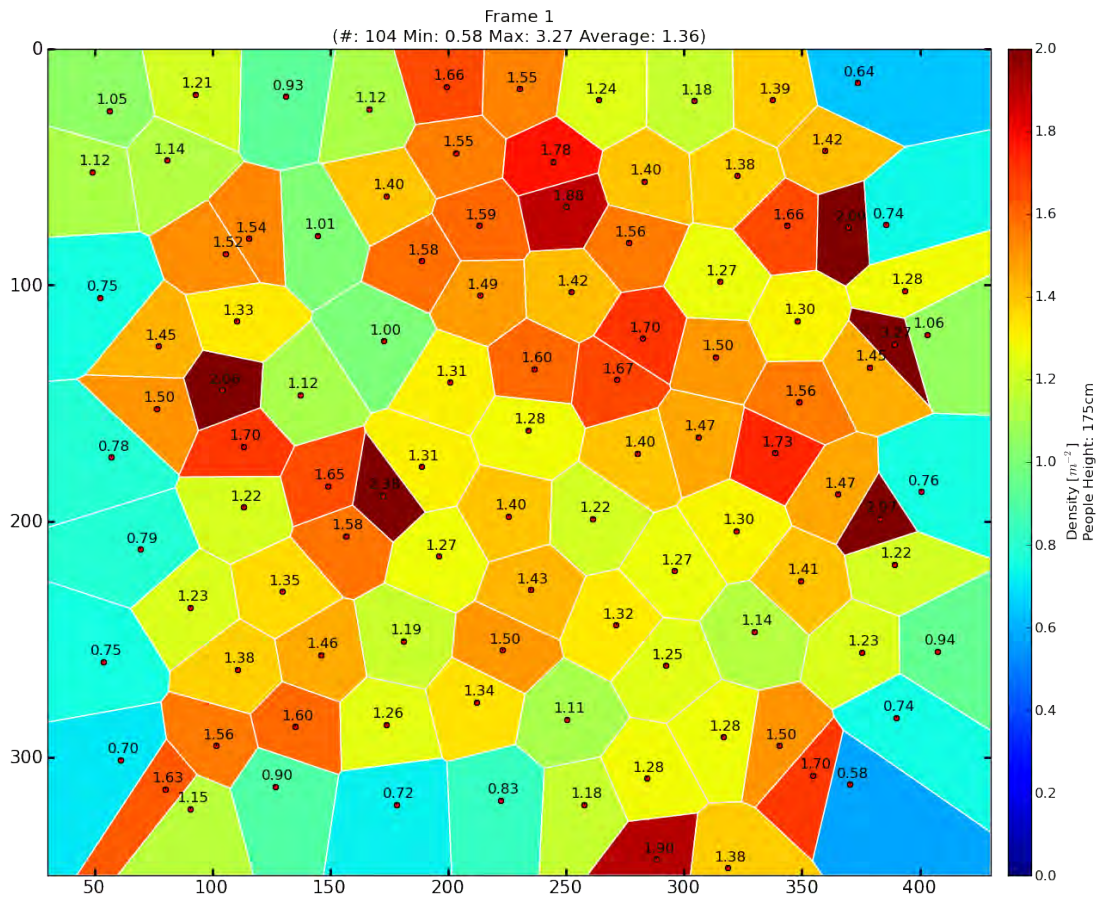
Der Flächeninhalt  $F_i$  ist umgekehrt proportional zur Dichte  $d_i$  in einer Zelle  $i$ . Je kleiner der Flächeninhalt, desto größer die Personendichte an dieser Stelle. Der Wert der Dichte für eine Person ist damit umgekehrt proportional zum Flächeninhalt seiner Voronoizelle.

$$d_i = \frac{1}{F_i}$$

Abbildung 4.2 zeigt das Voronoi-Diagramm einer Menschenmenge entsprechend ihrer Dichte eingefärbt. Die verwendeten Positionen sind dabei die auf den Boden projizierten Standpunkte der Personen [19, 20].

## 4.5 Dichtevisualisierung

Anhand der Dichtebestimmung im vorherigen Schritt erhalten wir für jede Person einen Wert mit der Dichte, welche die Person an dieser Position erfährt. Hohe Personendichten, in denen die Personen wenig Platz



**Abbildung 4.2:** Voronoi-Diagramm der Personendichten aufgrund der Standpositionen auf dem Boden

haben, werden dabei mit der Signalfarbe rot markiert. Weniger gefährliche Bereiche, in denen geringere Dichten vorliegen werden entsprechend grün bzw. blau markiert. Zu Grunde liegt dabei die übliche Farbskala von blau (tief) bis rot (hoch). Die eingeschriebene Zahl gibt die Dichte innerhalb der Zelle in Personen pro Quadratmeter an. Im Voronoidiagramm sind die berechneten Positionen für die Personen durch einen roten Punkt innerhalb der Zelle angegeben (vergleiche Abbildung 4.2).

Neben den Dichten für die einzelnen Personen wird außerdem die durchschnittliche Dichte der Personen im Bild sowie die maximale und die minimale Dichte berechnet.

Da nicht immer klar ist, ob die Zellen am Rand begrenzt sind oder nicht,

gibt es zwei Methoden mit den Randzellen umzugehen. Bei der ersten Methode wird ein rechteckiger Rahmen um die zu berechnenden Punkte gelegt. Die Voronoizellenberechnung bezieht dann den Rahmen als Begrenzung mit ein. Das Ergebnis zeigt Abbildung 4.2 auf der vorherigen Seite. Alternativ können die Voronoi-Zellen am Rand ignoriert werden. Dabei werden alle Zellen, die mindestens eine Seite am Rand des betrachteten Bereichs haben, nicht eingezeichnet. Ein Beispiel zeigt die Abbildung 5.6 auf Seite 47 von der Loveparade.

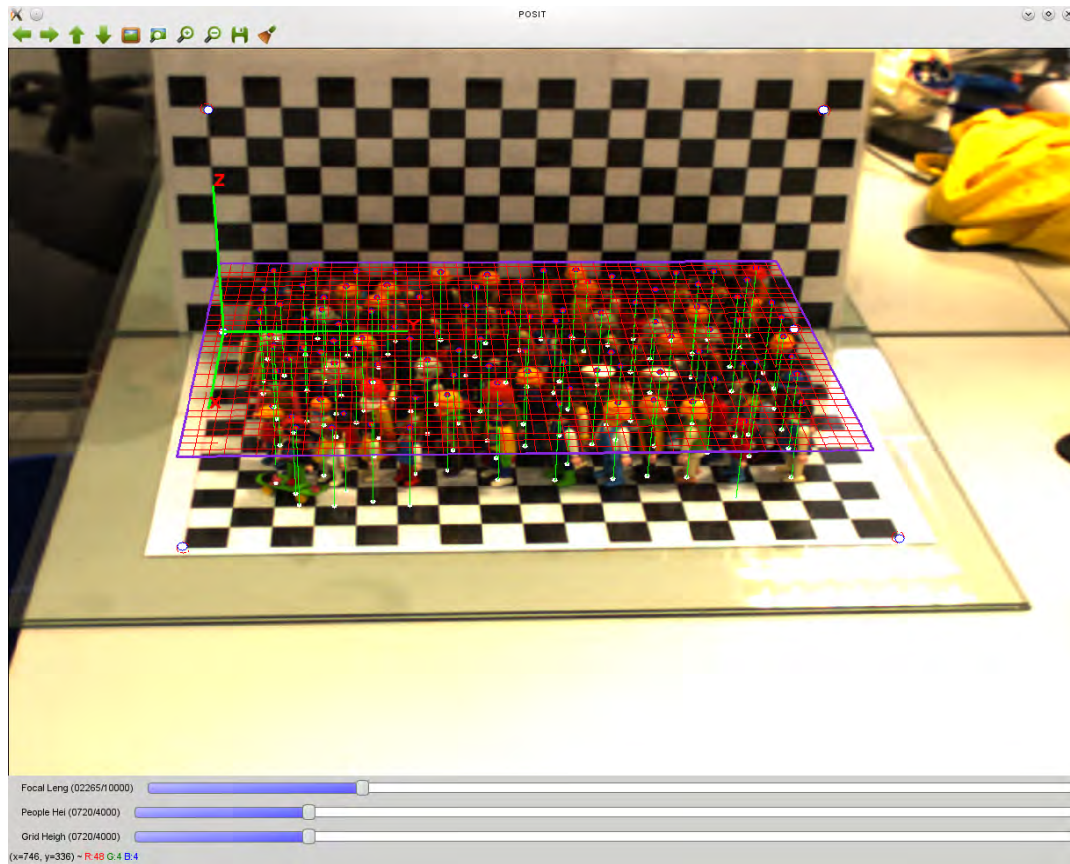
## 4.6 Grafische Oberfläche

Die Software wurde mithilfe der Programmierbibliothek OpenCV entwickelt. OpenCV ist eine freie unter der BSD-Lizenz<sup>1</sup> stehende C/C++-Programmierbibliothek, welche eine Fülle von Echtzeit-Funktionen zur Bildverarbeitung und Computer Vision beinhaltet. Sie bietet unter anderem bereits Funktionen zur Kamerakalibration und 3D-Rekonstruktion (Weitere Informationen zu OpenCV siehe Kapitel A.4 auf Seite 71).

Die grafische Oberfläche des Rahmenprogramms besteht aus einem Fenster, in dem das zu betrachtende Bild sowie Schieberegler zur Anpassung des Bildes zu sehen sind. Abbildung 4.3 zeigt einen Screenshot des Programms. Im unteren Bereich des Fensters sind Schieberegler platziert, mit denen man die eingestellte Brennweite, die angenommene Personengröße sowie die Höhe des eingeblendeten Gitters anpassen kann. Die Software bietet verschiedene ein- und ausblendbare Features an, wie z.B. das Weltkoordinatensystem, die Personenachse, das reguläre Gitter, die Voronoi-Polygone und die berechnete Position auf dem Boden (optional mit Koordinaten als Text). Neben der Bilddatei dienen als Eingabedaten Text-Dateien, in denen die 3D/2D-Punktkorrespondenzen als Kalibrationspunkte sowie die markierten 2D-Kopfpositionen angegeben werden. Die Parameter der Kamera-Matrix werden zu Beginn im Programm eingegeben.

---

<sup>1</sup> **Berkeley Software Distribution**, Software unter dieser Lizenz darf frei verwendet werden mit der Bedingung, dass der Copyright-Vermerk des ursprünglichen Programms nicht entfernt wird.



**Abbildung 4.3:** Zeigt die grafische Oberfläche der Software mit Schiebereglern und eingeblendeten Objekten

Als Ausgabe liefert das Programm die 2D-Standortpositionen der Personen auf dem Boden in einer Text-Datei. Außerdem kann die aktuelle Ansicht in der grafischen Oberfläche als Bilddatei abgespeichert werden. Die Voronoi-Zellen-Bestimmung sowie die Dichtebestimmung erfolgt mithilfe eines weiteren Moduls, welches in Kapitel A.2 auf Seite 64 zusammen mit einer Bedienungsanleitung ausführlich beschrieben wird.



# 5 Ergebnisse und Bewertung

## 5.1 Anwendung im Modell

Um die Güte des Verfahrens zu testen, wurde zunächst ein Modell mit Spielzeugfiguren aufgebaut. Die Spielzeugfiguren befinden sich auf einer ebenen Fläche vor einer Wand. Die Kamera wurde an einem Gestell frontal vor dem Modell befestigt und schaut auf die Spielzeugfiguren herab (vergleiche Abbildung 5.1 auf der nächsten Seite und die Aufbaubeschreibung des Modells in Kapitel A.1 auf Seite 63). Das Modell bietet den Vorteil, dass alle Positionen der Menschen exakt ausgemessen werden können. Außerdem sind die Eigenschaften der Kamera, mit der das Bild aufgenommen wurde, bekannt. Es können also die durch das Programm berechneten Resultate direkt mit den exakten Werten verglichen werden.

Das Schachbrettmuster auf dem Boden und auf der hinteren Wand dient lediglich dazu, bestimmte Punkte gut im Bild markieren zu können. So ist es z.B. einfacher die Punktkorrespondenzen, welche zur Kalibration des Verfahrens benötigt werden, anhand der Schachbrettecken anzugeben. Wie in der Verfahrensbeschreibung in Kapitel 4.1 auf Seite 31 bereits beschrieben, werden zu den im Modell ausgemessenen 3D-Punkten die 2D-Bildpunkte mittels eines Bildverarbeitungsprogramms angegeben. Unter Verwendung der Kameramatrix und den Punktkorrespondenzen wird die äußere Orientierung der Kamera berechnet. Mit der berechneten Rotationsmatrix und dem Translationsvektor kann man nun Punkte der 3-dimensionalen Szene ins Kamerakoordinatensystem überführen und anschließend die Positionen auf der Bildebene berechnen. Für die angegebenen Punktkorrespondenzen wurde dies in Abbildung 5.1 auf der nächsten Seite exemplarisch durchgeführt. Im Bild sieht man an den Ecken des





**Abbildung 5.1:** Modell mit Spielzeugfiguren, um das Verfahren testen zu können

Schachbrettmusters rote Kreise mit einem weißen Kreis im inneren. Der rote Kreis ist dabei exakt der im Bild gemessene Bildpunkt. Der weiße Punkt ist der mithilfe der Rotationsmatrix und des Translationsvektors sowie der Kameramatrix auf die Bildebene projizierte korrespondierende 3D-Punkt. Wie die annähernd gleiche Position der weißen Punkte und roten Kreise zeigen, hat die Kalibration der äußeren Kameraparameter vernünftige Werte geliefert.

Um die Position des Weltkoordinatensystems im Bild anzuzeigen, ist ein grünes Koordinatensystem entsprechend der Orientierung des Koordinatensystems eingezeichnet. Außerdem ist es nun möglich ein reguläres rechteckiges Gitter parallel zur Standfläche entsprechend der Perspektive in die 3D-Szene zu legen. Mithilfe des abgebildeten Gitters in Abbildung 5.1 kann man nun einen ersten Eindruck über die Dichteverteilung der abgebildeten Personen erhalten.

Abbildung 5.2 auf der nächsten Seite zeigt nun auch die Positionen der



**Abbildung 5.2:** Der weiße Punkt auf dem Boden spiegelt die Position der Person wieder. Die grüne Linie zeigt die Personenachse. Die rote Markierung zeigt zwei fehlerhafte Positionsrechnungen aufgrund der unterdurchschnittlich kleinen Personengröße (vergleiche Kapitel 5.3 auf Seite 52).

Personen auf dem Boden. Die violett gefärbten Punkte auf den Personenköpfen sind die vom Benutzer manuell ausgewählten Bildkoordinaten für die Kopfposition der Personen im Bild. Die grünen Verbindungslinien vom Kopfpunkt bis zum weißen Punkt auf dem Boden symbolisieren die Achsen der jeweiligen Personen. Die weißen Punkte, die in der Regel auf dem Boden platziert sind, geben die Standortpositionen der Personen wieder. Mithilfe dieser Standortpositionen ergibt sich schließlich das Voronoi-Diagramm in Abbildung 5.3 auf der nächsten Seite, das für jede Person im Bild die jeweilige Voronoi-Zelle mit der berechneten Dichte anzeigt.

Das Voronoidiagramm in Abbildung 5.3 auf der nächsten Seite zeigt die Dichteverteilung innerhalb der Menschenmenge. Es wurden 104 Personen vom Benutzer markiert. Außerdem hat der Benutzer für die Größe der



Personen den Wert: 175 cm angenommen. Man sieht, dass manche Per-

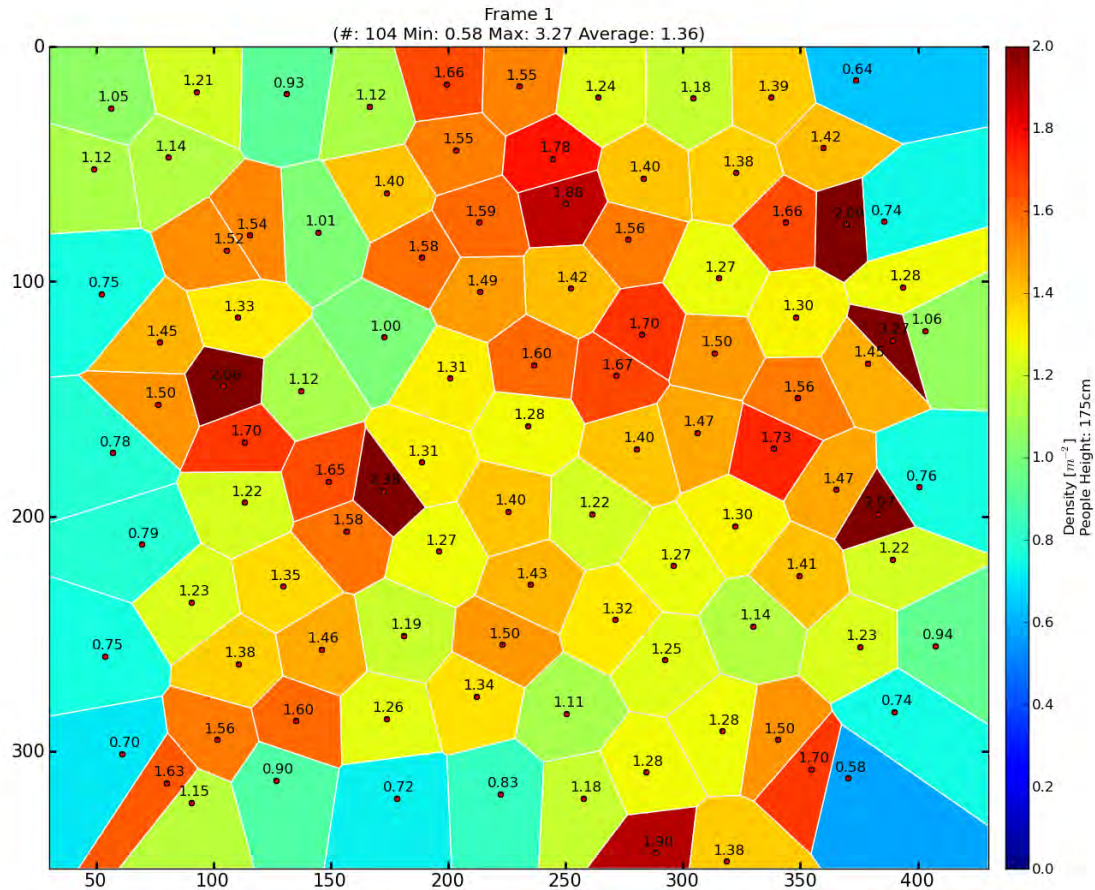


Abbildung 5.3: Voronoidiagramm der Personendichte

sonen, deren Voronoizellen dunkel rot bzw. braun gefärbt sind eine hohe Dichte von bis zu  $3.27 \frac{\text{Personen}}{m^2}$  erfahren. Andere Personen, die sich z.B. an den Rändern befinden erfahren eine geringere Dichte von weniger als einer Person pro  $m^2$ . Die Einheit  $\frac{\text{Personen}}{m^2}$  wird verwendet, da alle Angaben der 3D-Szene mit einem Faktor, der das Verhältnis zwischen Spielzeugfigurengröße und realer Menschengröße beinhaltet, skaliert wurden. Der Rand bei diesem Beispiel wurde hier durch einen Rahmen vorgegeben. Im Modell haben die Personen am rechten, linken und vorderen Rand keine Eingrenzungen zum Rand hin. Lediglich die Personen an der hinteren Wand werden durch diese entsprechend beeinträchtigt. Der Rand des Voronoi-Diagramms muss in der Regel separat betrachtet werden, da die Flächen der Randzellen keine Aussagekraft mehr über die erfahrende

Dichte haben.

Um die Voronoizellen direkt im Bild zu sehen, wurden in Abbildung 5.4 die Polygone der Voronoizellen direkt in die 3D-Szene auf Höhe der Personenköpfe eingezeichnet. In diesem Beispiel wurden die Randzellen nicht



Abbildung 5.4: Voronoi-Diagramm im Bild

berücksichtigt. Auch hier gewinnt man einen Eindruck davon, wie viel Platz die einzelnen Personen zur Verfügung haben.

## 5.2 Anwendung an realem Szenario

Um das implementierte Verfahren zur Dichtebestimmung von Menschenmassen an einem realen Szenario zu testen, wurde das Gelände der Loveparade 2010 in Duisburg ausgemessen und als Anwendung in einem realen Szenario verwendet. Als Ausgangsbild wurde ein Foto einer Über-



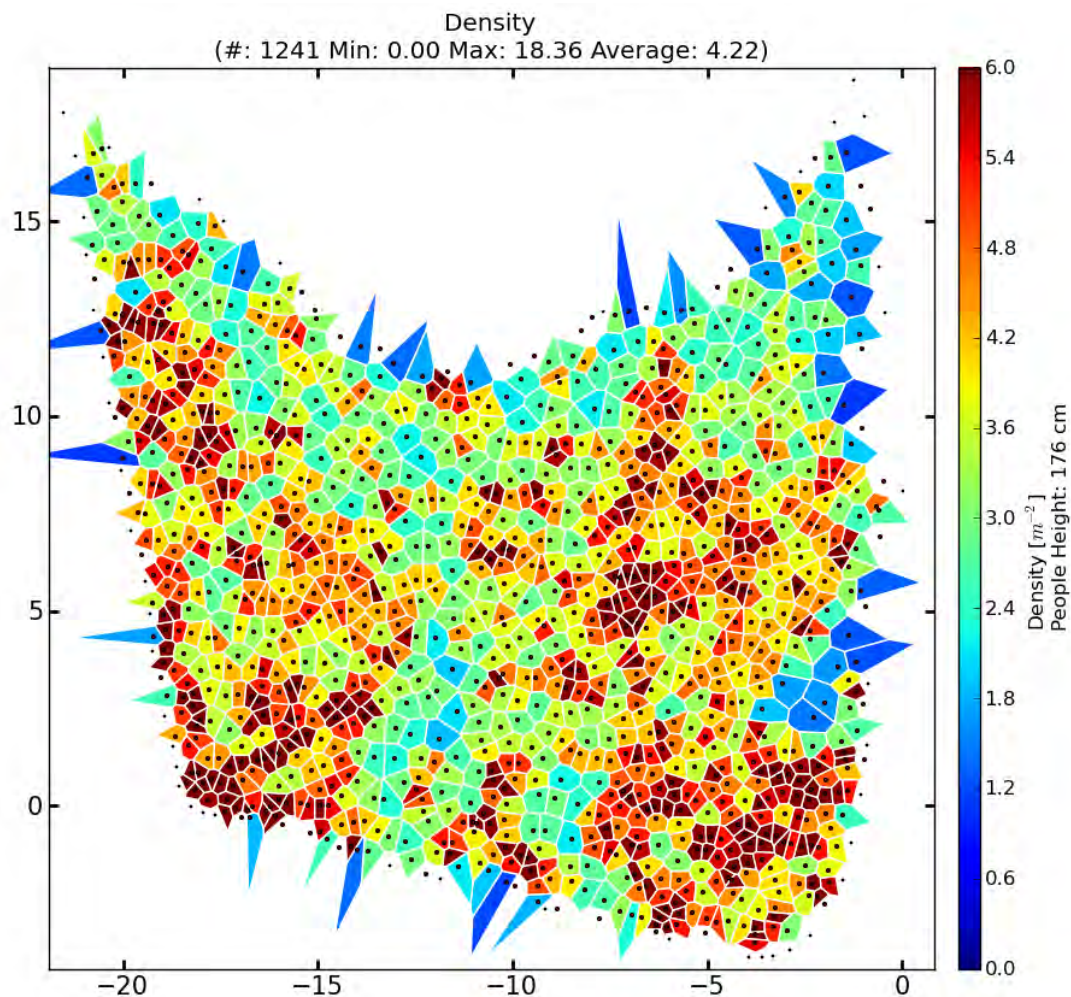
wachungskamera der Veranstaltung verwendet. Sämtliche Parameter der Kamera (Brennweite, Position, etc.) sind unbekannt und zur Berechnung der Dichte nicht verfügbar. Zunächst benötigt das Verfahren Punktkorrespondenzen aus der realen 3D-Szene und der 2D-Bildebene. Die 2D-Punkte werden mithilfe eines Bildverarbeitungsprogramms ausgewählt und notiert. Aufgrund der großen Besucherzahl zum Aufnahmezeitpunkt des Bildes ist es unmöglich Punkte auf dem Boden zu berücksichtigen. Dies hat den Nachteil, dass man keine Punkte der Ebene, auf der sich die Besucher bewegen, erhält. Diese Ebene ist besonders wichtig, da sich die zu berechnenden Positionen der Personen auf dieser Ebene befinden. Um trotzdem die Lage der Ebene des Bodens zu erhalten, wurde beim Vermessen der 3D-Punkte auch Punkte auf dem Boden mit ausgemessen. Die



**Abbildung 5.5:** Die weißen Punkte (grünen Linie) zeigen die berechneten Positionen der markierten Personen.

Punkte können zwar nicht zur Kalibration der äußeren Orientierung der Kamera verwendet werden, geben aber hinterher die Geometrie zwischen den Punkten auf dem Boden und den Punkten zur Kalibration wieder. Als Kalibrationspunkte wurden mehrere Punkte an der rechten Mauer und dem darauf montiertem Geländer verwendet. Zudem wurden zwei Punkte am Gebäude im hinteren Bereich des Geländes verwendet. Die Koordinaten der 3D-Punkte wurden Vorort ausgemessen.

Da die Brennweite unbekannt ist, musste die Brennweite experimentell durch Tests ermittelt werden. Dazu wurde der Kalibrationsschritt mit verschiedenen Brennweiten durchgeführt und geschaut bei welcher Brennweite die projizierten Punkte am besten auf die markierten passen. Das beste Ergebnis sieht man in Abbildung 5.5 auf der vorherigen Seite. Die Kalibrationspunkte (weiße Punkte in rotem Kreis) passen nicht exakt aufeinander. Dies liegt daran, dass die Kameraparameter nicht exakt bekannt sind und somit keine Entzerrung durchgeführt werden konnte. Das Ergeb-

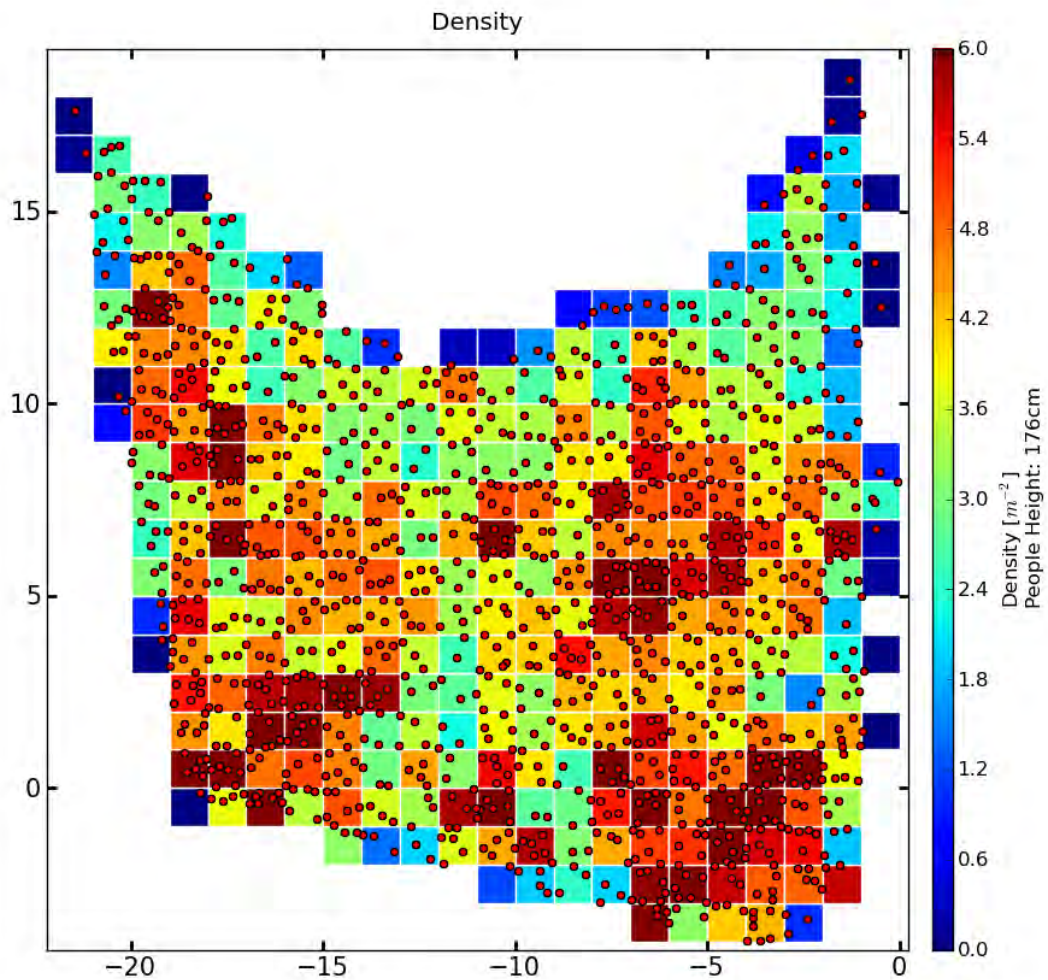


**Abbildung 5.6:** Voronoi-Diagramm der markierten Personen bei der Loveparade 2010 in Duisburg

nis der Voronoi-Zellenberechnung zeigt Abbildung 5.6. Man sieht deutliche Unterschiede in der Verteilung der Personen. Dabei herrscht eine durchschnittliche Dichte von  $4.22 \frac{\text{Personen}}{m^2}$ . Insgesamt wurden 1241 Personen



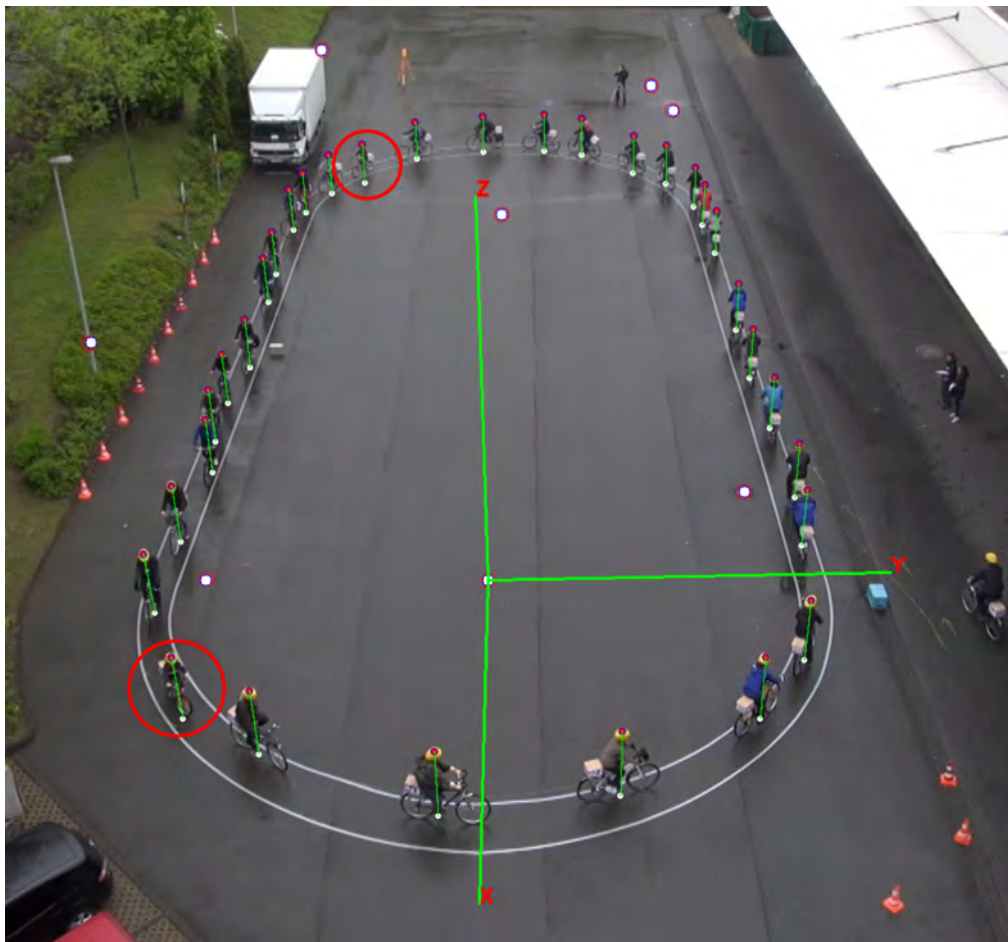
markiert, wobei es im Bild nicht einfach ist zu erkennen, was eine Person bzw. ein Kopf ist. Um weniger abhängig von der aktuellen Dichte der einzelnen Personen zu sein, zeigt Abbildung 5.7 die Dichteverteilung, bei der die Dichte in  $m^2$ -Blöcken angezeigt wird. Die Dichte für einen Quadratmeter berechnet sich dabei aus dem Durchschnitt der Dichten aller im Quadratmeter positionierten Personen. Das Diagramm bietet den Vorteil, dass extrem hohe Dichten einzelner Personen durch die umliegenden Personendichten gemittelt werden. Denn genau dies geschieht in der Realität durch Bewegen der Personen ebenso. Weiterhin rot markiert bleiben Bereiche, in denen mehrere Personen eine entsprechend hohe Dichte erfahren.



**Abbildung 5.7:** Diagramm der Dichteverteilung bei der Loveparade 2010 in Duisburg

### 5.2.1 Fahrrad-Experiment

Als weitere Anwendung des Verfahrens wurde ein Fahrradexperiment verwendet, bei dem ca. 30 Fahrradfahrer hintereinander entlang eines auf dem Boden markierten Ovals fahren. Die Fahrradfahrer wurden dabei von einer Kamera aus dem fünften Stock aufgenommen. Ziel bei dieser Anwendung ist es nicht, die Dichte für die einzelnen Fahrradfahrer zu ermitteln, sondern die genauen Positionen für die Fahrradfahrer auf dem Boden zu bestimmen. Eine Reihe exakt ausgemessener Punkte diente als Grundlage für die Punktkorrespondenzen, welche zur Orientierungsbestimmung benötigt werden. Die einzelnen Radfahrer wurden wieder jeweils am Kopf markiert, um ihre Position in der Bildebene zu erhalten.



**Abbildung 5.8:** Übersicht des Fahrrad-Experiments, bei dem nur die Positionen auf dem Boden berechnet werden sollten. Die roten Kreise markieren kleine Fahrradfahrer, bei denen die Positionsbestimmung fehlerhaft ist.

Wie in Abbildung 5.8 auf der vorherigen Seite zu erkennen, liefert die Kalibration in dieser Anwendung wieder gut brauchbare Werte. Das hängt damit zusammen, dass die Kalibrationspunkte zuvor exakt ausgemessen wurden und auch im Bild gut zu erkennen sind.

Wie in den Bildern zuvor sind die Standortpositionen der Radfahrer durch einen weißen Punkt auf dem Boden markiert. Da es sich bei den Fahrradfahrern um eine Mischung von Kindern und Erwachsenen handelt variieren die Höhen der Fahrradfahrer teilweise recht stark. Dies spiegelt sich auch im Ergebnis der Berechnung wieder. Für alle Fahrradfahrer wurde eine Durchschnittshöhe gewählt. Für die Kinder bedeutet dies, dass sie deutlich größer angenommen werden als sie in Wirklichkeit zusammen mit dem Fahrrad sind. Folglich liegt die berechnete Position nicht genau unter der Person, sondern etwas davor. Dies zeigt auch Abbildung 5.9, bei der die 2D-Standortpositionen der einzelnen Fahrradfahrer aufgetragen sind. Man sieht, dass die ovale Fahrbahn nicht auf allen Positionen exakt eingehalten wird. Besonders im oberen und unteren Halbkreis gibt es Abweichungen (rote Markierungen), die zum einen durch die unterschiedliche Fahrweise der Fahrer aber auch durch die Größenunterschiede der Personen hervorgerufen werden.

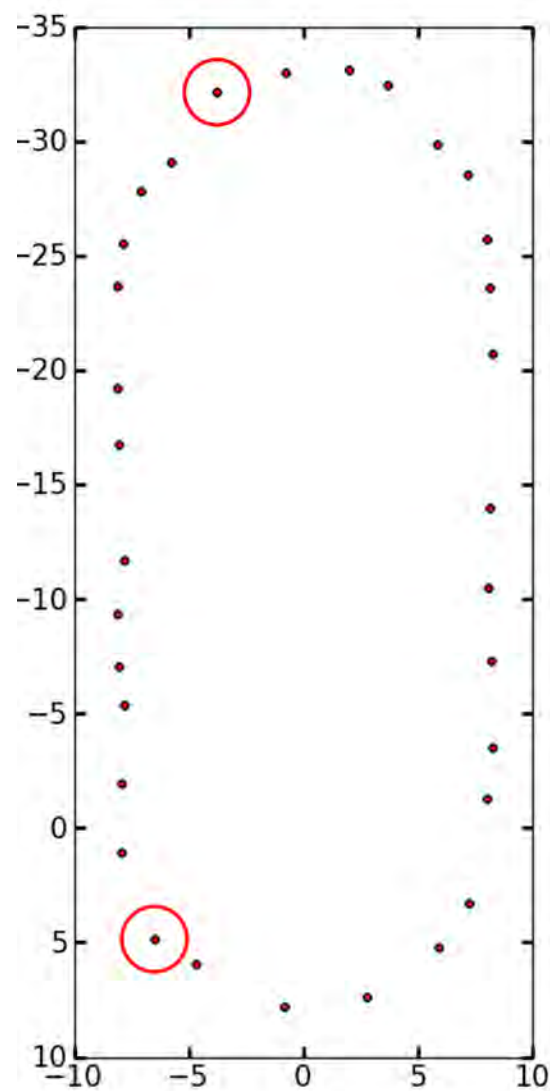
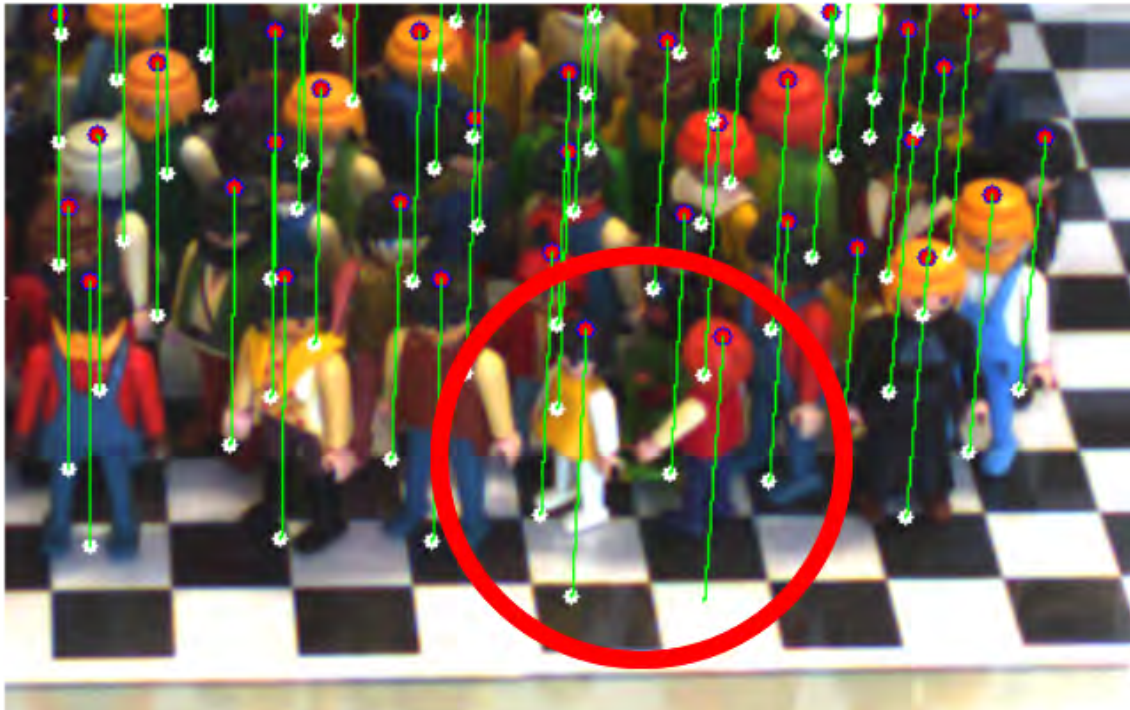


Abbildung 5.9: 2D-Standortpositionen der Fahrradfahrer



## 5.3 Fehlerbetrachtung

Da nicht für jede markierte Person die genau Körpergröße vorhanden ist, nimmt man bei dem Verfahren eine für alle Personen identische Durchschnittsgröße an. Folglich kommt es zu Fehlern bei der Positionsberechnung für die Personen, welche nicht die Durchschnittsgröße haben. Abbildung 5.10 zeigt zwei kleinere Personen, für die auch die Durchschnittsgröße angenommen wird. Folglich kommt es wie im Bild zu sehen zu einer zu weit zum Betrachter positionierten Positionsbestimmung. Je nachdem wie die Verteilung der übrigen Personen in der Umgebung der zu kleinen oder zu großen Personen ist, kommt es zu extrem hohen oder tiefen Dichten, welche jedoch in der Realität nicht vorliegen.

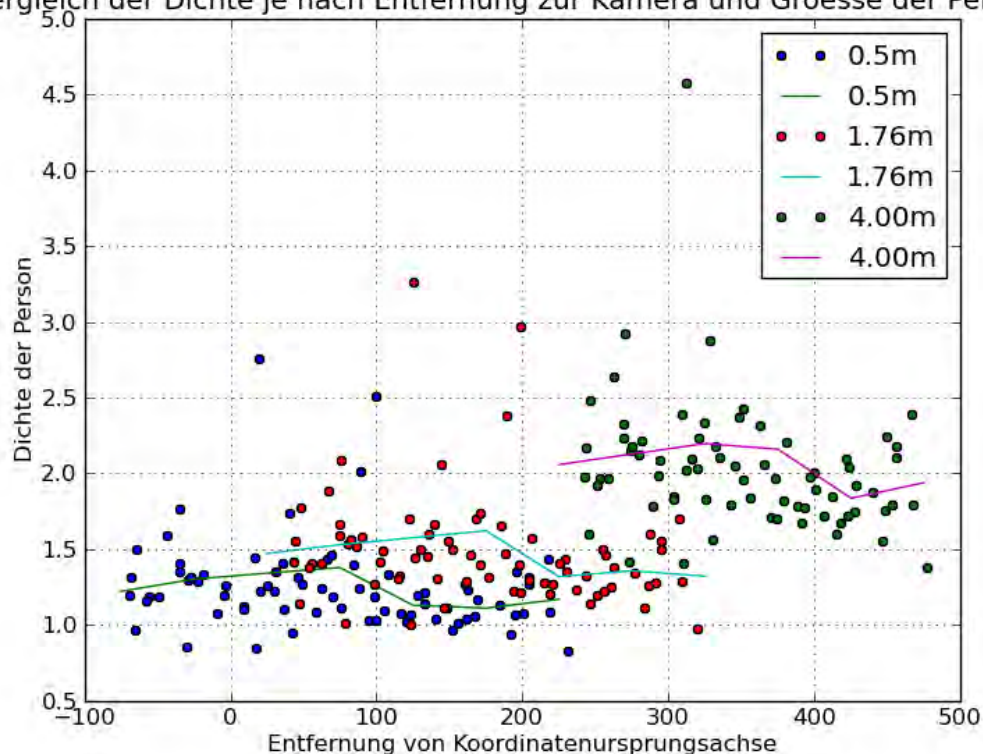


**Abbildung 5.10:** Die Positionen für unterdurchschnittlich kleine oder überdurchschnittlich große Personen sind fehlerhaft.

Außerdem soll in diesem Abschnitt untersucht werden, welche Auswirkungen eine größere oder eine kleinere angenommene durchschnittliche Personengröße auf das Ergebnis der Dichtebestimmung hat. Abbildung 5.11 auf der nächsten Seite zeigt den Vergleich von drei verschiedenen angenommenen Größen. Generell ist zu sehen, dass sich die durchschnittliche Dichte

mit der Höhe der angenommenen Personengröße auch vergrößert. Wenn man sich vorstellt, unter den markierten Punkten wären größere Personen als die Vorhandenen, so wären die Positionen näher zur Kamera und die Personen würde im Vergleich zur korrekten Größe deutlich enger stehen. Die Kopfpositionen der Personen würden sich näher zur Kamera befinden und somit wird der Bereich, in dem die Personen stehen ebenfalls kleiner. Umgekehrtes gilt für die zu klein angenommene Personengröße.

Vergleich der Dichte je nach Entfernung zur Kamera und Groesse der Personen

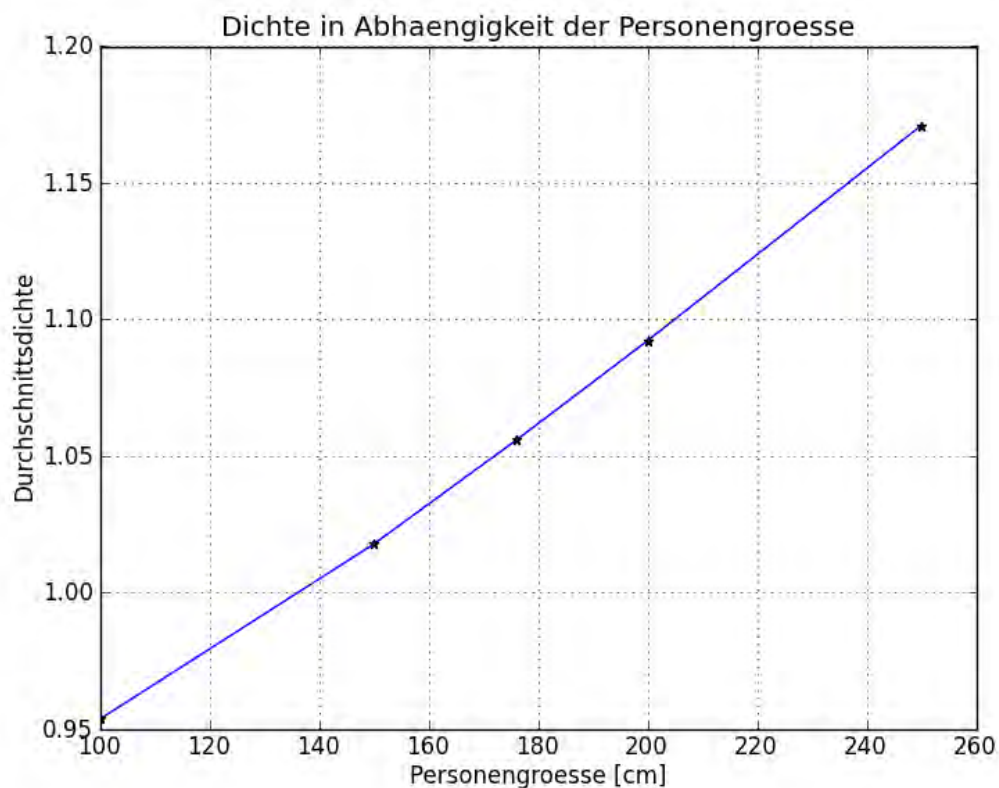


**Abbildung 5.11:** Vergleich von unterschiedlich angenommener Personengröße zur Dichte und zur Entfernung zur Kamera

Die X-Achse des Diagramms zeigt die Entfernung der Personen von der Y-Achse des Weltkoordinatensystems im Modell der Spielzeugfiguren. Je größer die Entfernung wird, desto näher kommt man der Kamera. Man sieht, dass größer angenommene Personen näher zur Kamera hin angezeigt werden. Kleiner angenommene Personen werden hingegen weiter entfernt angezeigt, zum Teil sogar im negativen Bereich hinter der Wand, die sich bei 0 befindet. Dieses Ergebnis ist auch nachvollziehbar, da man immer

von der markierten Position im 2D-Bild ausgeht. Stellt man sich nun vor, der gesamte Körper einer großen Person müsste nun noch unter dem markierten Punkt auf dem Boden stehen, so müsste die Person näher zur Kamera stehen. Umgekehrtes gilt wieder für kleinere Personen (siehe auch Kapitel A.3 auf Seite 67).

Die Abhängigkeit der Durchschnittsdichte von der angenommenen Personengröße spiegelt sich auch in Abbildung 5.12 wieder, welches die durchschnittliche Dichte gegen die angenommene Personengröße aufträgt. Es

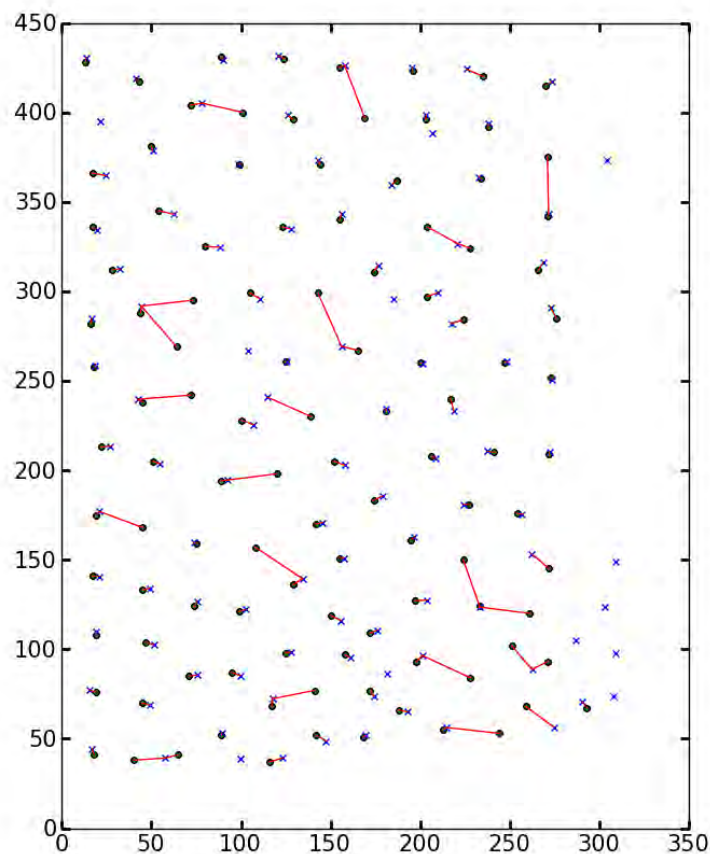


**Abbildung 5.12:** Abhängigkeit der Dichte von der Personengröße

lässt sich jedoch festhalten, dass bei einem realistischen Fehler bei der Annahme einer Durchschnittsgröße im Mittel von 10 cm sich die Durchschnittsdichte nicht signifikant verändert (vergleiche Abbildung A.3 auf Seite 70).

### Vergleich der berechneten Positionen mit den Originalpositionen

Um das Verfahren in seiner Güte bewerten zu können, wurde das Modell der Spielzeugfiguren genutzt, um die Abweichung der berechneten Positionen von den Originalpositionen zu betrachten. Abbildung 5.13 zeigt die Originalstandpunkte als grün-schwarzer Kreis und die ermittelten Positionen mit einem blauen Kreuz. Zusätzlich wurde für jeden Originalpunkt der nächstgelegene berechnete Punkt durch eine rote Verbindungslinie markiert. Bei den meisten Punkten ist die rote Linie fast nicht zu sehen,



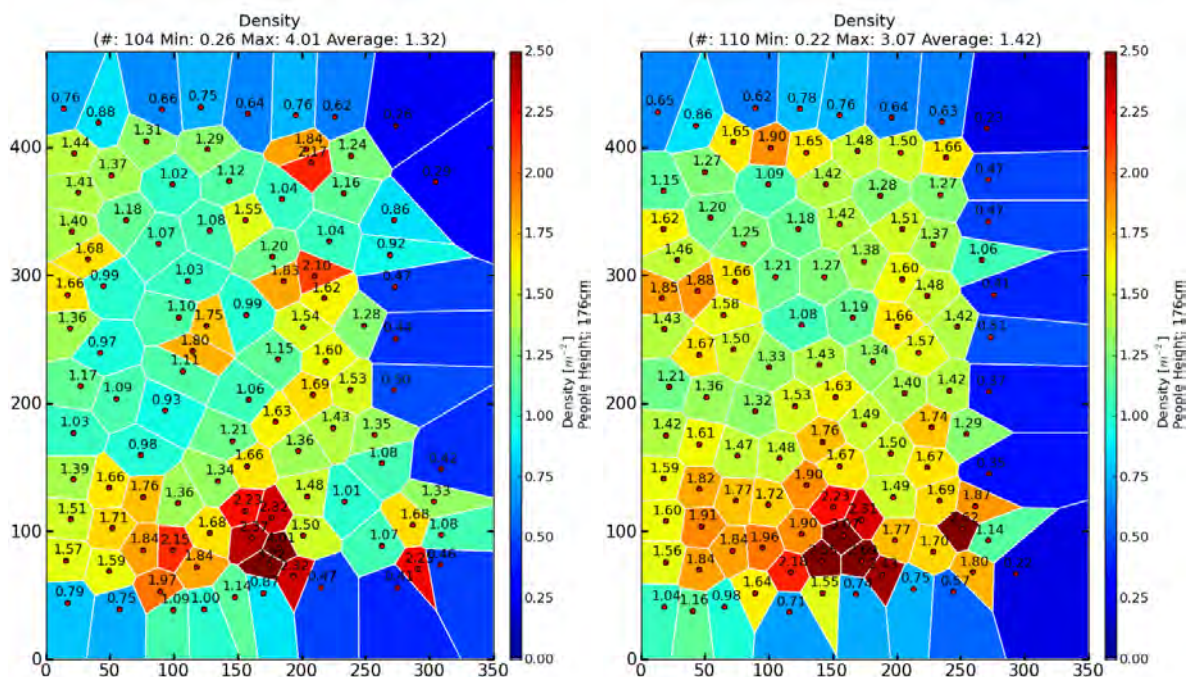
**Abbildung 5.13:** Vergleich der Originalstandpunkte (grün-schwarzer Kreis) und die berechneten Positionen (blaues Kreuz)

da die berechnete Position kaum abweicht. Das Auftreten einer langen roten Linie kann zwei Gründe haben. Zum Einen liegen die berechneten Positionen für kleinere Personen deutlich weiter vorne als der Originalpunkt. In dem Fall ist die nächstgelegene berechnete Position z.B. die



eines direkten Nachbarn in der Umgebung. In Folge dessen werden einer berechneten Position zwei Originalpunkte durch die rote Linie zugewiesen (vergleiche im unteren rechten Bereich des Bildes). Gleiches gilt bei größeren Personen, bei denen die berechnete Position weiter entfernt von der Kamera liegt. Der zweite Grund für eine lange rote Linie ist, wenn eine Person im Bild nicht markiert wurde. In dem Fall existiert zwar der Originalpunkt im Bild. Die berechnete Position ist jedoch nicht vorhanden, sodass die nächste Position wieder die des nächsten Nachbarn ist. Blaue Kreuze ohne rote Verbindungslinie zeigen entsprechend berechnete Punkte, die zu weit vom Originalpunkt abweichen.

Die Abweichung zwischen den Originalpunkten und den berechneten Punkten macht sich auch im Voronoi-Diagramm zu den Punkten deutlich. Abbildung 5.14 zeigt links das Voronoi-Diagramm der berechneten Positionen und rechts das Diagramm der original ausgemessenen Positionen. Es ist zu sehen, dass an den Positionen, wo kleinere Personen stehen, das Diagramm auseinander gezogen wird (vergleiche rechts unten im Bild). Von den 110 aufgestellten Personen wurden im Bild nur 104 markiert.



**Abbildung 5.14:** Vergleich der Voronoi-Diagramme der berechneten Positionen (links) und der Originalstandpunkte (rechts)

Bei den 6 fehlenden Personen handelt es sich um im Bild verdeckte Personen oder um Personen, die durch die mangelnde Qualität des Bildes nicht erkannt wurden. Im Regelfall ist die berechnete Durchschnittsdichte etwas zu tief, da im Bild verdeckte Personen nicht mit berücksichtigt werden.

Insgesamt bleibt festzuhalten, dass die Dichteverteilung nicht gravierend vom tatsächlichen Wert abweicht. Insbesondere, wenn von einer größeren Menschenmenge ausgegangen wird. Da sich die Menschen ständig bewegen, ist die Dichte für eine Person zu einem bestimmten Zeitpunkt nicht so relevant, wie die Gesamtverteilung der Personen im Bild bzw. die Durchschnittsdichte für einen bestimmten Bereich im Bild.



## 6 Zusammenfassung

Ziel der Arbeit war es, die Personendichte für eine aufgenommene Menschenmenge unabhängig von der Kameraausrichtung zu ermitteln. Die Arbeit hat gezeigt, dass anhand des beschriebenen Verfahrens eine Dichtebestimmung unabhängig von der Kameraausrichtung möglich ist. Vorausgesetzt man verfügt über mindestens 4 Punktkorrespondenzen von 2D-Bildpunkten und 3D-Weltkoordinaten sowie den Parametern der Kameramatrix. Die Bestimmung der Rotation und Translation vom Weltkoordinatensystem ins Kamerakoordinatensystem wurde anhand des POSIT-Verfahrens ermittelt. Nachdem man die äußere Orientierung der Kamera berechnet hat, ist man in der Lage, Aussagen über die perspektivische Verzerrung im Bild zu treffen. Man kann z.B. ein Rechteck oder eine Linie perspektivisch in die 3D-Welt legen. Zur Berechnung der Personenpositionen wird eine Schnittpunktberechnung zwischen der Geraden vom Projektionsursprung durch den 2D-Bildpunkt in die Szene und der Ebene der Personenköpfe durchgeführt. Der Schnittpunkt liefert schließlich den 3D-Punkt für den Kopf der Person. Der Standpunkt befindet sich senkrecht unter dem Kopf auf dem Boden. Die Berechnung der Dichte erfolgt anschließend mithilfe des Voronoi-Diagramms. Der Flächeninhalt der Voronoi-Zelle ist dabei umgekehrt proportional zur Dichte der Person.

Das Verfahren wurde an einem Modell und an zwei realen Szenarien getestet. Mithilfe des Modells, bei dem die Menschenmenge durch Spielzeugfiguren nachgebildet ist, besteht die Möglichkeit, das Verfahren in seiner Güte zu testen, da die Kameraposition und die Verteilung der Personen bekannt ist. Das Verfahren liefert brauchbare Ergebnisse. Lediglich die Unterschiede in den Personengrößen führen bei einer festangenenen Durchschnittsgröße zu Fehlern bei besonders kleinen und besonders



großen Personen. In einem für ein Fahrrad-Experiment aufgebautem Szenario wurde das Verfahren verwendet, um die Positionen der Radfahrer auf dem Boden zu bestimmen. Bei der Anwendung auf das Bild der Loveparade ergaben sich kleinere Probleme, welche jedoch nicht direkt mit dem Verfahren sondern mit den äußeren Umständen zu tun haben. So ist es nicht einfach aufgrund der mäßigen Bildqualität die Positionen der Personenköpfe zu markieren. Es werden zum Teil schnell Rucksäcke als Köpfe markiert oder manche Köpfe sind im Bild nicht als Köpfe zu erkennen. Außerdem waren die Parameter der Kameramatrix völlig unbekannt und mussten experimentell ermittelt bzw. geschätzt werden. Trotz alledem ist auch das Ergebnis des Loveparadebildes durchaus akzeptabel und verwendbar.

Desweiteren wurden Untersuchungen zur Güte des Verfahrens in Abhängigkeit von der angenommenen Personengröße sowie der Position der Person im Bild durchgeführt.

## 6.1 Ausblick

Für den Live-Einsatz ist das Verfahren erst tauglich, wenn man das Finden bzw. Markieren der Personen in einem automatischen Prozess durchführen kann. Erst dann kann das Verfahren auch als Live-Anwendung verwendet werden, welches parallel zu einer Veranstaltung die Besuchermassen analysiert.

Für ein Warnsystem bei einer Veranstaltung wäre es hilfreich Bereiche, in denen eine durchschnittlich hohe Personendichte vorliegt, bereits frühzeitig zu finden. Bisher erhält man Dichteangaben nur zu einzelnen Personen nicht jedoch zu einem weitergefassten Bereich. Von Interesse wäre jedoch auch, zu untersuchen, ob in einem bestimmten Bereich durchgehend eine hohe Personendichte vorliegt. Somit wäre dieser Bereich besonders gefährdet. Bereiche in denen eine Person extrem dicht steht aber alle weiteren umliegenden Personen ausreichend Platz haben, würden so nicht kritisch gesehen, da die Person im Normalfall auch nach kurzer Zeit wieder mehr Platz hat.

Das Einblenden des Voronoi-Diagramms im Bild der grafischen Oberfläche entsprechend der Perspektive wird zurzeit ohne die dichteabhängige Einfärbung der Voronoi-Zellen durchgeführt. Hier könnte man durch halbtransparente Farben die Dichte direkt im Bild entsprechend verdeutlichen. Das hätte den Vorteil, dass man die einzelnen Zellen direkt anhand der Farbe vergleichen könnte. Schließlich werden durch die Perspektive gleich große Zellen, die unterschiedlich weit von der Kamera entfernt liegen, unterschiedlich groß angezeigt.

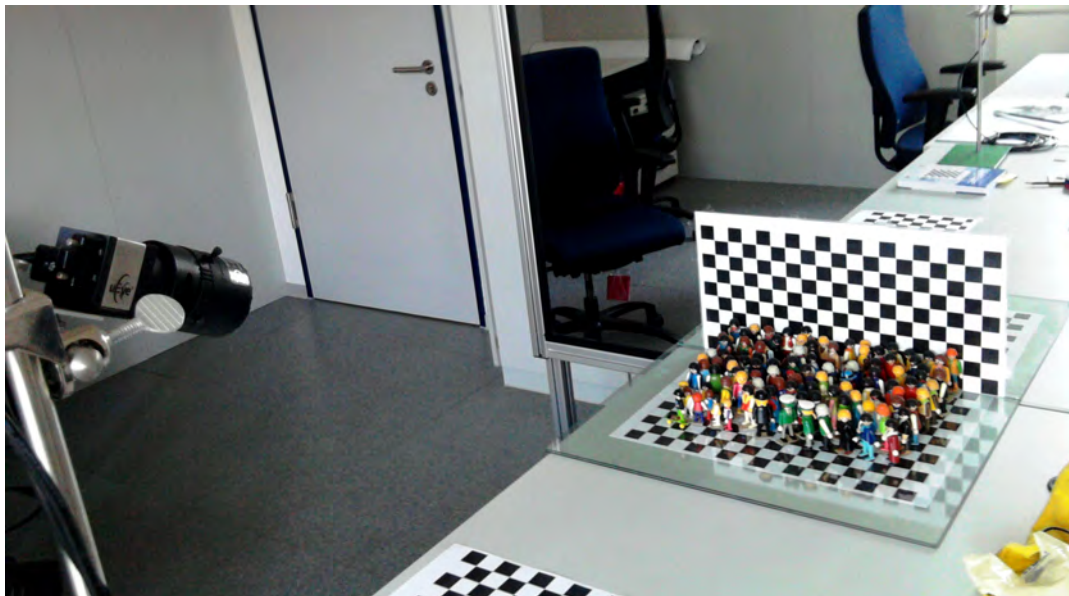
Für ein breiteres Anwendungsspektrum wäre es interessant die Positions- und Dichtebestimmung unabhängig von der Annahme, dass sich alle Personen auf einer Ebene befinden, zu betrachten. So wäre es denkbar, die Geometrie des Bodens händisch anzugeben oder das Gelände vorher durch einen flächendeckenden 3D-Scanner aufzunehmen. Schließlich schränkt die Annahme, dass sich alle Personen auf einer ebenen Fläche befinden, das Verfahren in seiner praxistauglichkeit deutlich ein.



# A Anhang

## A.1 Modellaufbau

Den Aufbau des in Kapitel 5.1 auf Seite 41 betrachteten Spielzeugfiguren Modells zeigt Abbildung A.1. Wie beschrieben blickt die Kamera frontal von einem höheren Punkt auf die Spielzeugfiguren hinunter. Der Ursprung des Kamerakoordinatensystems befindet sich im Projektionsursprung der Kamera. Die optische Achse der Kamera verläuft in Richtung der Menschenmenge. Das Weltkoordinatensystem befindet sich in der hinteren linken Ecke der Menschenmenge, wie in Abbildung 5.1 auf Seite 42 grün eingezeichnet.



**Abbildung A.1:** Modellaufbau des Spielzeugfiguren Modells

## A.2 Benutzungsanleitung

Die Software besteht aus zwei eigenen Bestandteilen, sowie einem externen Programm zum berechnen der Voronoi-Zellen. Der erste Teil der Software kümmert sich um die Positionsberechnung der markierten Personen. Im zweiten Teil werden die Voronoi-Zellen Polygone weiterverarbeitet zu einem Voronoi-Diagramm. Bevor man die Software startet benötigt man folgende Voraussetzungen:

### Voraussetzungen

***Entzerrte Bilddatei*** Das zu untersuchende Bild muss bereits entzerrt sein. Die Verzerrungsparameter sowie die Entzerrung werden in Kapitel 2.5 auf Seite 18 genauer beschrieben.

***3D-Kalibrierungspunkte*** Zur Angabe der 3D-Koordinaten der Kalibrationspunkte muss eine Text-Datei erstellt werden. In der ersten Zeile steht die Anzahl der Punkte, darunter folgen die Punkte zeilenweise mit X-, Y- und Z-Wert.

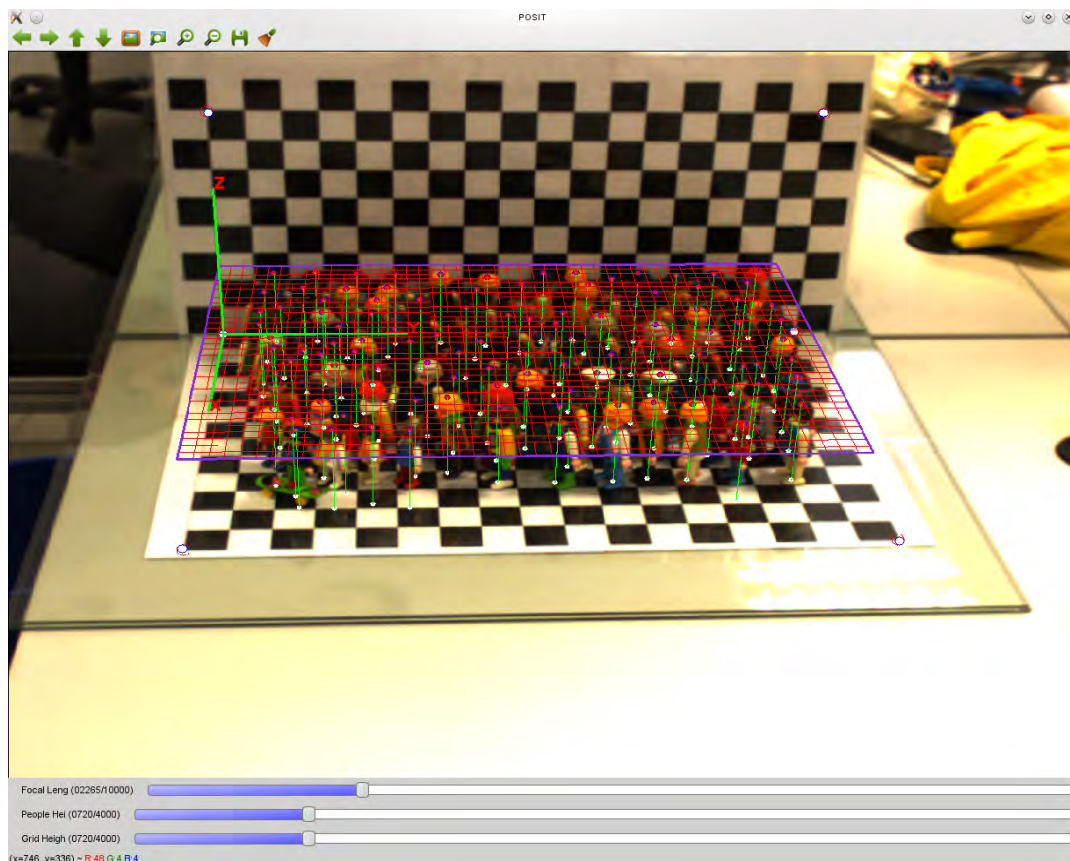
***2D-Kalibrierungspunkte und 2D-Personenpunkte*** Die Angabe der 2D-Bildkoordinaten erfolgt auch per Text-Datei. Zunächst wird die Anzahl, anschließend die Koordinaten der Kalibrierungspunkte selber, angegeben. Hinter den Kalibrierungspunkten folgen die Koordinaten der Personen nach dem selben Muster.

***Elemente der Kameramatrix*** Die Elemente der Kameramatrix werden einmal zu Beginn im Programm eingegeben.

Die Software bietet verschiedene einblendbare Objekte an. Folgende Auflistung bietet eine Übersicht über die verfügbaren Optionen, welche jeweils über Schieberegler oder Boolean-Werte ein- oder ausgeblendet werden können.

- 3D-Weltkoordinatensystem
- 2D-Bildkoordinatensystem
- Reguläres gleichmäßiges Gitter entsprechend der Perspektive

- Berechnete Positionen auf dem Boden (optional mit Koordinaten als Text)
- Personenachse
- Voronoi-Polygone der Personen
- OpenCV eigene Zoom-Funktion



**Abbildung A.2:** Zeigt die grafische Oberfläche der Software mit Schiebereglern und eingeblendeten Objekten

Abbildung A.2 zeigt die grafische Oberfläche der Software. Im oberen Bereich des Fensters sind OpenCV eigene Funktionen wie die Zoom- oder Speicher-Funktion platziert. Im unteren Bereich des Fensters befinden sich Schieberegler zur Steuerung des Programms. Im Fenster ist das zu untersuchende Bild mit 3D-Weltkoordinatensystem (grün), ein rechteckiges reguläres Gitter (rot) sowie die markierten (violett) und berechneten Positionen (weiss + grüne Linie) der Personen markiert.

Neben den über die Benutzungsoberfläche einstellbaren Optionen bietet das Programm auch die Möglichkeit die berechneten Positionskoordinaten als Text-Datei abzuspeichern. Ebenso kann das im Fenster angezeigte Bild mit allen angezeigten Objekten im Bild als Bilddatei abgespeichert werden. Die Software besitzt außerdem Schieberegler, um die Brennweite, die angenommene Personengröße sowie die Höhe des eingeblendeten regulären Gitters zu verändern (vergleiche Abbildung A.2 auf der vorherigen Seite).

Über die als Text-Datei abgespeicherten Positionskoordinaten erhält man nun vom externen Programm die einzelnen Eckpunkte der Voronoi-Zellen. Der zweite Bestandteil der Software berechnet zu den Voronoi-Zellen nun die Flächen und daraus die Dichte für die einzelnen Personen. Durch einzeichnen aller Voronoi-Zellen entsteht das Voronoi-Diagramm, welches im Laufe der Arbeit genauer beschrieben wurde. Um die Ausgabe nach den eigenen Wünschen anzupassen existieren folgende Einstellungsmöglichkeiten:

- Randzellen ein-/ausblenden
- Dichte als Text ein-/ausblenden
- Anpassung der Farbskala
- Positionspunkte ein-/ausblenden
- Dichte pro Voronoi-Zelle oder pro  $m^2$  anzeigen
- Angaben zum Bild (Anzahl Punkte, min., max. und durchschnittliche Dichte) ein-/ausblenden

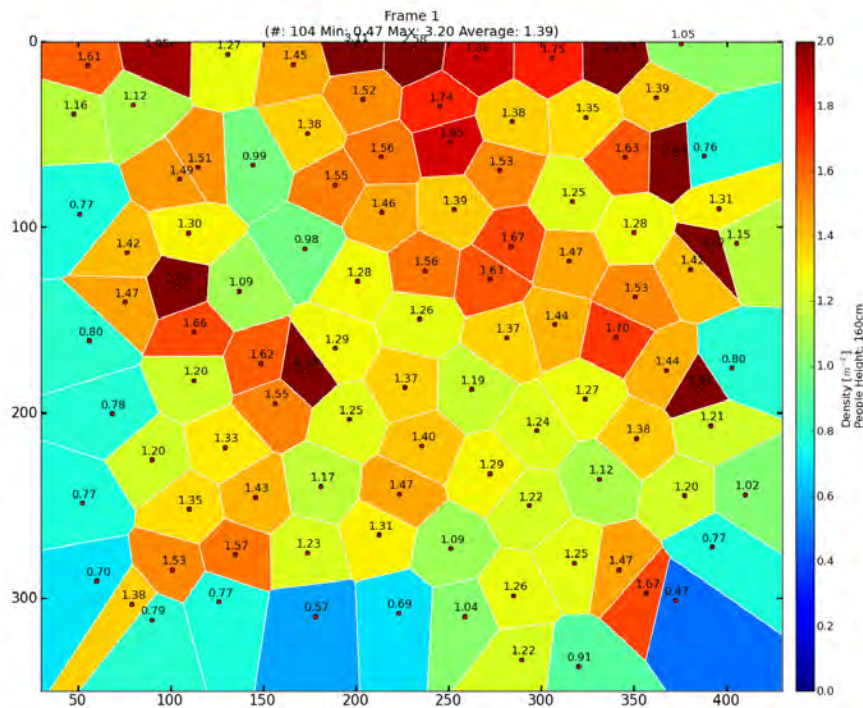
Die Einstellungen werden dabei über **Boolean**-Werte im Programm vorgenommen:

```
#Options
show_points = True
show_edges = True
show_bound_cells = True
show_point_density = True
show_patch = True
```

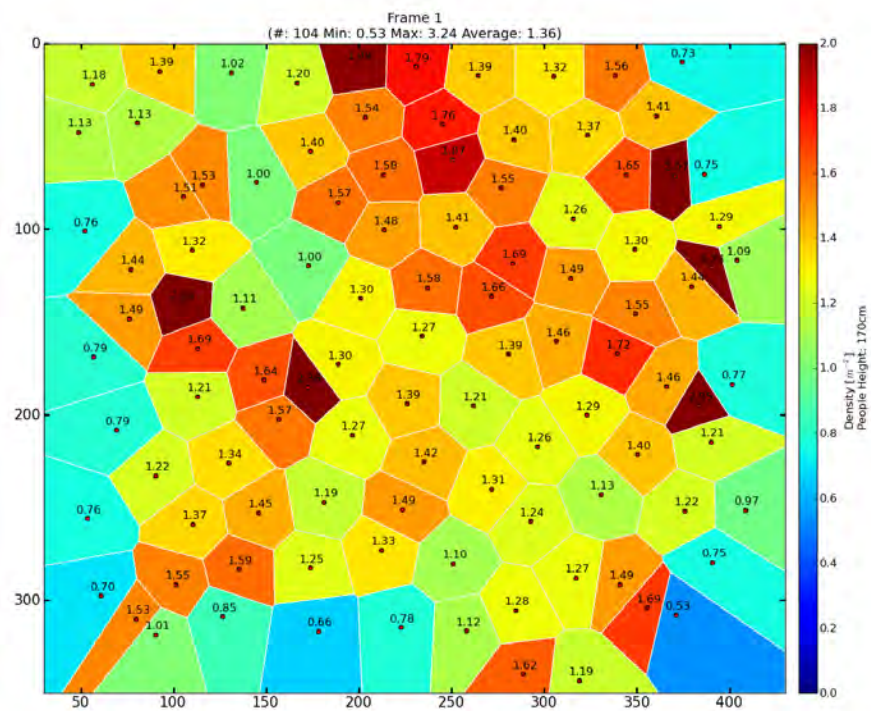
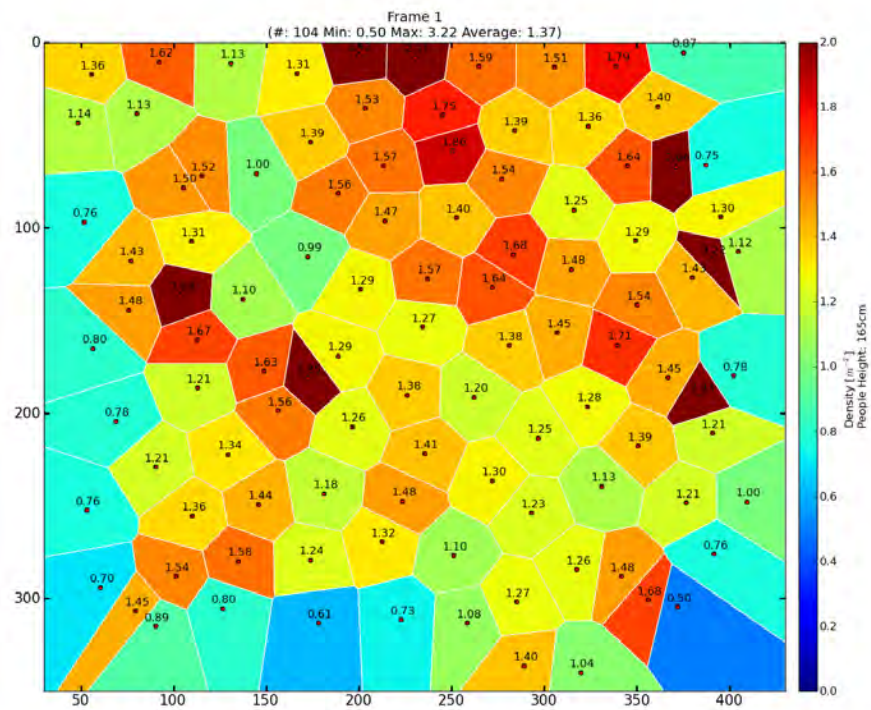
Da es sich bei der Software um Python einer Interpreter Sprache handelt, können Änderungen im Programmcode ohne großen Aufwand vorgenommen werden.

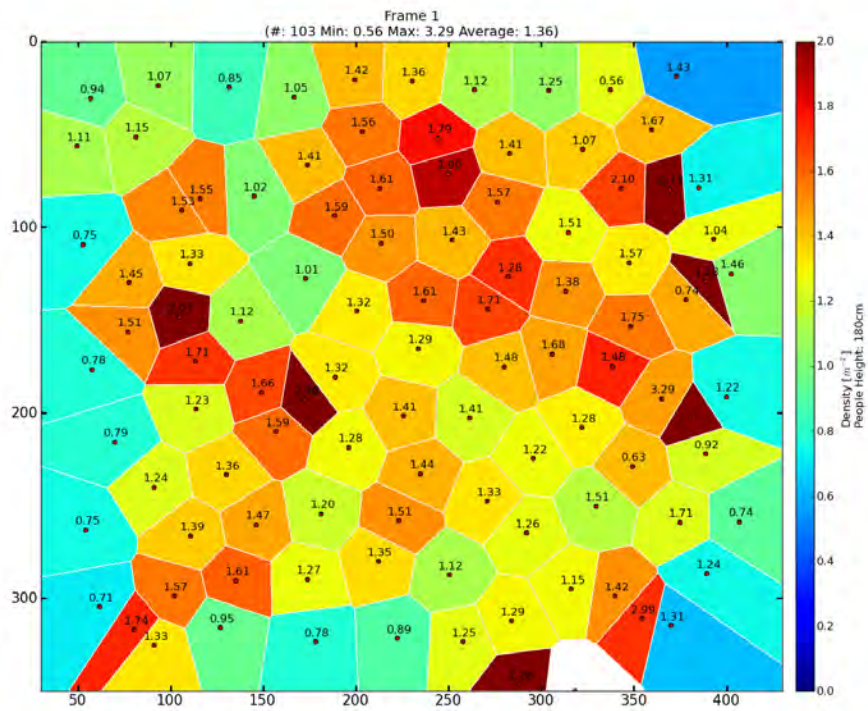
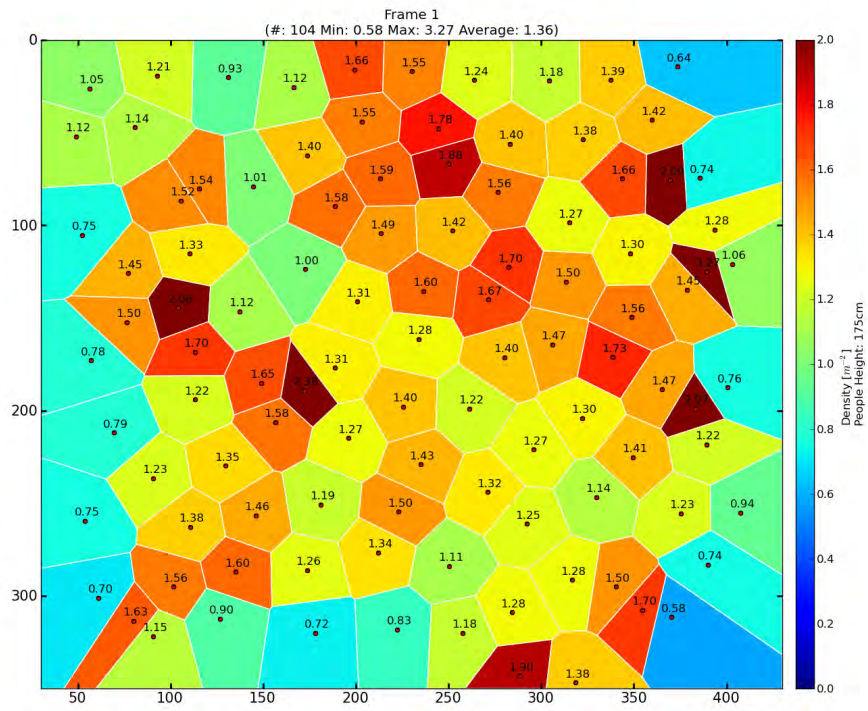
## A.3 Ergebnisse verschiedener angenommener Größen

Die folgenden Abbildungen (A.3) zeigen die Ergebnisse der Dichteverteilung für das Spielzeugfiguren-Modell, wobei als angenommene Personengröße die Größen von 1,60 m, 1,65 m, ... bis 1,90 m untersucht wurden. Die Abbildungen verdeutlichen die Verschiebung der Personenpositionen je nach angenommener Personengröße.

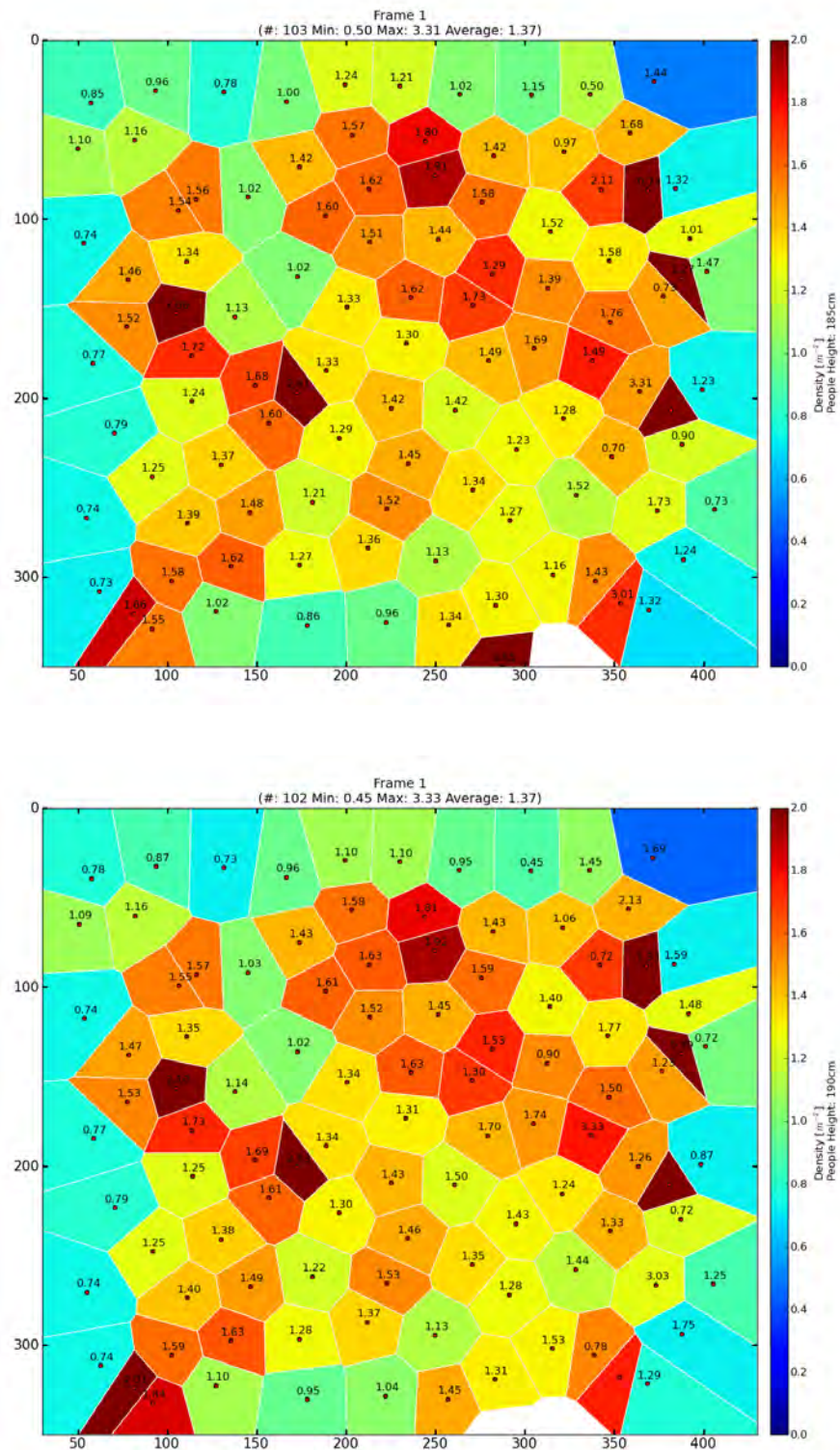












**Abbildung A.3:** Verschiebung der Positionskoordinaten und Dichteänderungen auf Grund der unterschiedlich angenommenen Personengröße

## A.4 OpenCV

OpenCV ist eine freie unter der BSD-Lizenz stehende C/C++-Programmierbibliothek, welche eine Fülle von Echtzeit-Funktionen zur Bildverarbeitung und Computer Vision beinhaltet. Mit mehr als 2500 Algorithmen zur Bildverarbeitung und zum maschinellen Sehen liegt sie derzeit in der Version 2.4.1 (Juni 2012) vor. Entwickelt wurde die OpenCV-Bibliothek für die Plattformen: Windows, Linux, Mac und Android. Die Version 1.0 wurde 2006 von Intel herausgegeben. Mittlerweile wird die Bibliothek jedoch weitestgehend von Willow Garage, einem US-amerikanischen Unternehmen für Robotertechnik, weiterentwickelt. Den Quellcode findet man auf [sourceforge.net](http://sourceforge.net), wo er permanent mit den neuesten Ergebnissen der Forschung von einer großen Benutzergemeinde erweitert wird. Die Bibliothek ist in mehrere Module unterteilt, die sich z.B. mit Objekterkennung, 3D Kalibrierung oder Bildfiltern befassen. Zur Implementierung des in dieser Arbeit beschriebenen Verfahrens wurden hauptsächlich Funktionen des Moduls Kamera Kalibrierung und 3D-Rekonstruktion verwendet [9, 13].

## A.5 Verwendete Hilfsmittel

- Qt Software Development Kit (QtSDK) Version 1.2
- Qt Creator IDE Version 2.4.1
- Programmier-Bibliotheken
  - OpenCV (Open Source Computer Vision Library) Version 2.3
  - Boost C++-Library Version 1.50.0
- Verwendete Literatur siehe Literaturverzeichnis



# Abbildungsverzeichnis

2.1	Zusammenhang der Koordinatensysteme . . . . .	6
2.2	Lochkameramodell . . . . .	11
2.3	Vergleich Zentralprojektion/Parallelprojektion . . . . .	12
2.4	Externe Transformation . . . . .	14
2.5	Perspektivische Projektion . . . . .	15
2.6	Radiale Verzerrung . . . . .	19
2.7	Übersicht Transformationsschritte . . . . .	20
3.1	Kamerakalibrierungsmuster . . . . .	22
3.2	Vergleich der perspektivischen und der skalierten senkrechten Projektion . . . . .	24
4.1	Schnittpunktberechnung Gerade-Ebene . . . . .	34
4.2	Voronoi-Diagramm der Personendichten aufgrund der Standpositionen auf dem Boden . . . . .	37
4.3	Grafische Oberfläche der Software . . . . .	39
5.1	Spielzeugmodell . . . . .	42
5.2	Spielzeugmodell mit berechneter Position . . . . .	43
5.3	Voronoi-Diagramm für das Spielzeugmodell . . . . .	44
5.4	Voronoi-Diagramm im Bild . . . . .	45
5.5	Loveparadebild mit berechneten Positionen . . . . .	46
5.6	Voronoi-Diagramm zum Loveparadebild . . . . .	47
5.7	Dichteverteilung zum Loveparadebild . . . . .	48
5.8	Übersicht des Fahrrad-Experiments . . . . .	49
5.9	Berechnete Positionen der Fahrradfahrer . . . . .	51
5.10	Fehlerbetrachtung bei kleinen/großen Personen . . . . .	52
5.11	Abhängigkeit der Dichte/Verteilung jenach angenommener Personengröße . . . . .	53
5.12	Abhängigkeit der Dichte von angenommener Personengröße . . . . .	54

---

5.13	Vergleich der Originalstandpunkte mit den berechneten Standpunkten . . . . .	55
5.14	Vergleich der Voronoi-Diagramme von den Originalstandpunkten und den berechneten Standpunkten . . . . .	56
A.1	Modellaufbau . . . . .	63
A.2	Grafische Oberfläche der Software . . . . .	65
A.3	Vergleich der Dichte für verschiedene Größen . . . . .	70

# Literaturverzeichnis

- [1] VDI Technologiezentrum, *Bausteine für die Sicherheit von Großveranstaltungen*. Bundesministerium für Bildung und Forschung (BMBF), <http://www.bmbf.de/pubRD/BaSiGo.pdf> (Aufgerufen: 26.07.2012), 2012
- [2] Prof. Dr. Armin Seyfried (JSC), *Hermes - Erforschung eines Evakuierungsassistenten für den Kriesenfall bei Großveranstaltungen*. Bundesministerium für Bildung und Forschung (BMBF), [http://www.bmbf.de/pubRD/SuRvM\\_Bekanntm.\\_600x800\\_D\\_-HERMES.pdf](http://www.bmbf.de/pubRD/SuRvM_Bekanntm._600x800_D_-HERMES.pdf) (Aufgerufen: 13.08.2012), 2012
- [3] Mikel Rodriguez, Ivan Laptev, Josef Sivic, Jean-Yves Audibert, *Density-aware person detection and tracking in crowds*, Imagine, LIGM, Universite Paris-Est, 2011
- [4] Mikel Rodriguez, Josef Sivic, Ivan Laptev, Jean-Yves Audibert, *Data-driven Crowd Analysis in Videos*. Imagine, LIGM, Universite Paris-Est, 2011
- [5] Wikipedia.org, *Lochkamera*. <http://www.de.wikipedia.org/wiki/Lochkamera> (Aufgerufen 30.07.2012), 2012
- [6] S. Graf, *Kamerakalibrierung mit radialer Verzeichnung. Die radiale essentielle Matrix*. Universität Passau, 2007
- [7] Daniel F. DeMenthon, Larry S. Davis, *Model-Based Object Pose in 25 Lines of Code*. Computer Vision Laboratory Center for Automation Research, University of Maryland College Park, 1995
- [8] Sebastian Grembowietz, Prof. Gudrun Klinker, Prof. Dr. Nassir Navab *Algorithms for Augmented Reality - 3D Pose Estimation*. Technische Universität München (TUM), 2004



- [9] Gary Bradski, Adrian Kaehler *Learning OpenCV - Computer Vision with the OpenCV Library*. O'Reilly Media, First Edition, 2008
- [10] Dr. Oliver Schreer, *Stereoanalyse und Bildsynthese*. Fraunhofer Institut für Nachrichtentechnik, Heinrich-Hertz-Institut, 2005
- [11] Wikipedia.org, *Zentralprojektion*. <http://www.de.wikipedia.org/wiki/Zentralprojektion> (Aufgerufen: 03.07.2012), 2012
- [12] Wikipedia.org, *Strahlensatz*. <http://www.de.wikipedia.org/wiki/Strahlensatz> (Aufgerufen: 01.08.2012), 2012
- [13] OpenCV Development Team, *OpenCV v2.4.2 Dokumentation und C++-Referenz*. <http://www.docs.opencv.org> (Aufgerufen: 03.07.2012), 2012
- [14] Javier Barandiaran, *POSIT Tutorial - Pose Estimation*. OpenCV-Wiki: <http://www.opencv.willowgarage.com/wiki/Posit> (Aufgerufen: 03.07.2012), 2010
- [15] Wikipedia.org, *Koordinatentransformation*. <http://de.wikipedia.org/wiki/Koordinatentransformation>, (Aufgerufen: 04.07.2012), 2012
- [16] Wikipedia.org, *Projektionsmatrix*. <http://de.wikipedia.org/wiki/Projektionsmatrix> (Aufgerufen: 26.07.2012), 2012
- [17] Yihong Wu, Zhanyi Hu, *PnP Problem Revisited*. Journal of Mathematical Imaging and Vision 24:131-141, DOI: 10.1007/s10851-005-3617-z, 2006
- [18] Prof. Gudrun Klinker, Prof. Dr. Nassir Navab, T. Sielhorst, Ilina Dimitrova *Kamera Kalibrierung, Augmented Reality*. Technische Universität München (TUM), 2004
- [19] Wikipedia.org, *Voronoi-Interpolation*, <http://de.wikipedia.org/wiki/Voronoi-Interpolation> (Aufgerufen: 25.07.2012), 2012
- [20] Bernhard Steffen, Armin Seyfried, *Methods for measuring pedestrian density, flow, speed, and direction with minimal scatter*. Physica A (Seite 1902-1910), DOI: 10.1016/j.physa.2009.12.015, 2009

- [21] OpenCV.org (WillowGarage), *Camera Calibration and 3D Reconstruction*. [http://www.opencv.willowgarage.com/documentation/cpp/camera\\_calibration\\_and\\_3d\\_reconstruction.html](http://www.opencv.willowgarage.com/documentation/cpp/camera_calibration_and_3d_reconstruction.html) (Aufgerufen: 26.07.2012), 2012
- [22] Jörg Rädler, *A python package that handles polygonal shapes in 2D*. <http://www.j-raedler.de/projects/polygon/> (Aufgerufen: 25.07.2012), 2012
- [23] John Hunter, Darren Dale, Michael Droettboom, *Python 2D plotting library which produces publication quality figures*. <http://matplotlib.sourceforge.net/> (Aufgerufen 25.07.2012), 2008
- [24] Wikipedia.org, *Singular Value Decomposition*. [http://en.wikipedia.org/wiki/Singular\\_value\\_decomposition](http://en.wikipedia.org/wiki/Singular_value_decomposition) (Aufgerufen: 26.07.2012), 2012
- [25] Michael E. Wall, Andreas Rechtsteiner, Luis M. Rocha, *Singular value decomposition and principal component analysis*. Modeling, Algorithms, and Informatics Group (CCS-3), Los Alamos National Laboratory, New Mexico, USA, 2003



# Glossar

BaSiGo	Forschungsprojekt u.a. am Forschungszentrum Jülich mit dem Titel: ‘BaSiGo - Bausteine für die Sicherheit von Großveranstaltungen’. Ziel des Projektes ist es Großveranstaltungen sicherer zu machen. Da zu werden Menschenströme und -ansammlungen untersucht., 1
Bildebene	Dient als Projektionsfläche der Kamera. Die Bildebene liegt parallel zur $XY$ -Ebene des Kamerakoordinatensystems mit einem Abstand $f$ der Brennweite., 7
Bildhauptpunkt	Definiert den Schnittpunkt der optischen Achse mit der Bildebene., 22
Brennweite	Abstand zwischen der Bildebene und dem Projektionsursprung, 21
Dichtebestimmung	Beschreibt das Verfahren zur Bestimmung der Dichte für die einzelnen Personen im Bild., 32
Extrinsische Parameter	Beschreiben die äußere Orientierung der Kamera. Die Orientierung wird durch die Rotation und die Translation zum Weltkoordinatensystem angegeben., 21
Homogene Erweiterung	Fügt den Vektoren und Matrizen eine zusätzliche Zeile hinzu, um eine geschlossene Matrix-Schreibweise für alle Operationen zu ermöglichen., 9

Innere Orientierung	Beschreibt die interne Geometrie der Kamera. Dabei werden die Kameraspezifischen Eigenschaften in der Kameramatrix angegeben., 21
Intrinsische Parameter	Beschreiben die innere Geometrie einer Kamera. Dabei wird die Kameramatrix mit Brennweite, Bildhauptpunkt und Skalierungsfaktor angegeben. Zusätzlich können die Verzerrungsparameter des Kameramodell angegeben werden., 21
Kamerakalibration	Beschreibt das Verfahren zur Ermittlung der äußeren und inneren Orientierung der Kamera., 21
Kamerakoordinatensystem	Definiert das Koordinatensystem der Kamera., 6
Kameraparameter	Beschreiben die internen Eigenschaften einer Kamera, wie Brennweite, Bildhauptpunkt und Verzerrungskoeffizienten, 18
Lochkamera	Beschreibt ein einfaches Modell, bei dem die 3D-Szene auf eine 2D-Projektionsfläche abgebildet wird., 10
OpenCV	C/C++-Programmierbibliothek für den Bereich der Bildverarbeitung und Computer Grafik., 22
Perspektivische Projektion	Projektion, bei der sich die Projektionsstrahlen in einem gemeinsamen Projektionsursprung schneiden. Nähere Objekte werden größer abgebildet, als gleichgroße Objekte weiter entfernt., 12
Pixelskalierung	Beschreibt das Verhältnis von Pixel zu Meter., 21

PnP (Perspective-n-Point Problem)	Beschreibt das Problem anhand $n$ Punktkorrespondenzen die äußere Kameraorientierung zu ermitteln., 22
POSIT	Verfahren zum Lösen des PnP-Problems., 23
Positionsbestimmung	Beschreibt das Verfahren zur Standortpositionsbestimmung für die einzelnen Personen in einem aufgenommenen Bild, 32
Radiale Verzerrung	Beschreibt die zum Linsenmittelpunkt hin radial abnehmende kissen- oder tonnenförmige Verzerrung., 18
Rotationsmatrix	Mathematische Beschreibung einer Drehung durch eine Matrix., 14
Singulärwertzerlegung	Verfahren zum Lösen eines Linearen Gleichungssystems, 28
Skalierte Senkrechte Projektion (SOP)	Verfahren zur Approximation der perspektivischen Projektion, bei dem zunächst eine orthogonale Projektion auf eine Ebene parallel zur Bildebene und anschließend die Ebene perspektivisch auf die Bildebene projiziert wird., 25
Tangentiale Verzerrung	Beschreibt die tangential am Kreis durch den Punkt anliegende Verzerrung, aufgrund eines Konstruktionsfehlers der Kamera., 19
Translationsvektor	Mathematische Beschreibung einer Verschiebung durch einen Vektor, 14
Voronoi-Diagramm	Beschreibt das Diagramm, in dem alle Voronoizellen eingezeichnet sind., 36
Voronoi-Zelle	Beschreibt die Zelle um eine Person, in der jeder Punkt der Zelle näher zur eigenen Person als zu allen anderen Personen ist., 36

Weltkoordinatensystem	Beschreibt das Referenzkoordinatensystem des 3D-Raums. Es ist der Ursprung für alle Koordinaten im 3-dimensionalen Raum., 5
-----------------------	---

# Index

- Äußere Orientierung, *siehe* Extrinsische Parameter
- BaSiGo, 1
- Bildebene, 7
- Bildhauptpunkt, 22
- Brennweite, 21
- Dichtebestimmung, 32, 36
- Extrinsische Parameter, 21
- Homogene Erweiterung, 9
- Innere Orientierung, *siehe* Intrinsische Parameter
- Intrinsische Parameter, 21
- Kamerakalibration, 22
- Kamerakoordinatensystem, 7
- Kameraparameter, 18
- Lochkamera, 10
- OpenCV, 22
- Perspektivische Projektion, 12
- Pixelskalierung, 21
- PnP, 23
- POSIT, 23
- Positionsberechnung, 33
- Positionsbestimmung, 32
- Radiale Verzerrung, 18
- Rotationsmatrix, 14
- Singulärwertzerlegung, 28
- Skalierte Senkrechte Projektion (SOP), 26
- Tangentiale Verzerrung, 19
- Translationsvektor, 14
- Voronoi-Diagramm, 36
- Voronoi-Zelle, 36
- Weltkoordinatensystem, 5



Jül-4355  
September 2012  
ISSN 0944-2952